

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 11, 2005, 09:31:07 ; Search time 3864 Seconds
(without alignments)
3856.105 Million cell updates/sec

Title: US-09-522-753-5
Perfect score: 13215
Sequence: 1 MSGSTQLVAVQWTRATEPRYP.....WDEPRKLLCSQYETLSDSE 2517

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=x1d
-Q=/cgn2_1/USPTO.epool_p/US09522733/runat_02112005_173640_20785/app_query.faeta_1.2695
-DB=N.GeneSeq.16Dec04 -QFMT=faetap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=b1ts -START=1 -END=-1 -MATRIX=b1osum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCL -OUTFMT=p2n -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09522733@cgn_1_1_1431@runat_02112005_173640_20785 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N.GeneSeq.16Dec04:*
1: geneSeqn1980s:*
2: geneSeqn1990s:*
3: geneSeqn2000s:*
4: geneSeqn2001as:*
5: geneSeqn2001bs:*
6: geneSeqn2002as:*
7: geneSeqn2002bs:*
8: geneSeqn2003as:*
9: geneSeqn2003bs:*
10: geneSeqn2003cs:*
11: geneSeqn2003ds:*
12: geneSeqn2004as:*
13: geneSeqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13215	100.0	7554	12	ADJ92815 Human CO-
2	13215	100.0	8561	12	ADG86298 Human SMR
3	13215	100.0	8561	12	ADN04304 Human SMR
4	13178.5	99.7	8564	3	AACT4783 Human ORF
5	13145	99.5	8667	11	ACN44283 Human mRN

6	13119.5	99.3	8533	13	ADG84524 Human tum
7	13119.5	99.3	8533	13	ACN93603 Tumour-as
8	12978	98.2	7521	8	ACA62250 Human nuc
9	12978	98.2	7524	10	ADL13812 Osteoarth
10	12978	98.2	8686	6	ACA62249 cDNA enco
11	12978	98.2	8686	10	ADL13811 Osteoarth
12	12978	98.2	8686	12	ADG86290 Human SMR
13	12978	98.2	8686	12	ADQ18920 Human sof
14	12922	97.8	9053	12	ADL12577 Human ste
15	12845.5	97.2	9079	12	ADQ23294 Human sof
16	10832.5	82.0	7386	8	ACA62452 Mouse nuc
17	10832.5	82.0	8544	8	ACA62451 cDNA enco
18	9691.5	73.3	7534	11	ACN44281 Mouse mRN
19	7885	59.7	5989	6	ABK84305 Human CDN
20	4612	34.9	2930	10	ADG35130 Human bre
21	4612	34.9	2930	12	ADG86301 Human SMR
22	4187	31.7	7780	3	AA60630 HNRCR nuc
23	4165.5	31.5	7914	12	ADG83959 Human tum
24	4165.5	31.5	7914	13	ADG87683 Human tum
25	4147.5	31.4	7940	8	AB234833 Coding se
26	4147.5	31.4	7940	13	ADG89791 Antagonis
27	4043	30.6	7900	3	AA60629 Human HNR
28	3649	27.6	9141	11	ACN44280 Mouse gen
29	2951	22.3	23380	11	ACN44282 Human gen
30	2184	16.5	2745	12	ADJ92816 Human co-
31	2133.5	16.1	220756	12	ADG86300 Human SMR
32	1871.5	14.2	2336	10	AD631306 Human dia
33	1391	10.5	2914	13	ADR07585 Full leng
34	1113	8.4	650	8	ACA57401 Human adi
35	1099	8.3	10410	13	ADG56633 Drosophi1
36	1083	8.2	10910	4	ABL03131 Drosophi1
37	1076	8.1	872	3	AA602670 Human col
38	1073.5	8.1	10900	12	ADG89593 Antagonis
39	977.5	7.4	2519	13	ADP48942 Human-yea
40	960	7.3	555	8	ACA57524 Human adi
41	880	6.7	718	10	AD676358 Human BSK
42	845	5.9	673	8	ACA57523 Human neu
43	774.5	5.6	520	6	ABK44934 cDNA enco
44	746.5	5.5	33529	5	AA617367 DNA seque
45	726.5	5.5	33529	5	AA617367 DNA seque

ALIGNMENTS

RESULT 1	ADJ92815	ADJ92815 standard; DNA; 7554 BP.
XX	ADJ92815;	
AC	ADJ92815;	
XX		
DT	06-MAY-2004 (first entry)	
XX		
DE	Human co-repressor SMRT DNA.	
XX		
KW	Nuclear receptor; therapy; diabetes; hyperlipidaemia; obesity;	
KM	atherosclerosis; human; co-repressor; gene; ds.	
XX		
OS	Homo sapiens.	
XX		
PN	US2003228607-A1.	
XX		
PD	11-DEC-2003.	
XX		
PF	14-APR-2003; 2003US-00414692.	
XX		
PR	15-APR-2002; 2002US-0372650P.	
XX		
PA	(WAGN/) WAGNER B L.	
XX	(SCHU/) SCHULMAN I G.	
PI	Wagner BL, Schulman IG;	
XX	WPI; 2004-167207/16.	
DR		

XX Identifying compounds that bind to nuclear receptor and exhibit cell type
PT specific actions, and useful for treating hyperlipidemia, obesity and
PT diabetes.
XX
XX Claim 5; SEQ ID NO 14; 99pp; English.
XX
XX The invention relates to screening methods for identifying compounds that
CC bind to nuclear receptor and exhibit cell type specific actions. The
CC invention relates to modulators having an improved therapeutic profile.
CC The method is useful for identifying compounds that bind to a nuclear
CC receptor and exhibit cell type specific actions. It is also useful for
CC identifying modulators of nuclear receptors that are useful in treating
CC diseases e.g., diabetes, hyperlipidaemia, obesity, atherosclerosis, etc.
CC The present sequence is human co-repressor DNA used to illustrate the
CC method of the invention.

SO Sequence 7554 BP; 1646 A; 2732 C; 2228 G; 948 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0	Length:	7554
Score:	13215.00	Matches:	2517
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	12	Gaps:	0

US-09-522-753-5 (1-2517) x ADJ92815 (1-7554)

QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTPArgAlaThrGluProArgTyrPro 20
DB 1 AAGTCGGGCTCCACACAGCTTGTCGACAGACGTGAGGGCCACTGAGCCCGCTACCGG 60
QY 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeu 40
DB 61 CCCACAGCCTTTCCTACCAAGTCAGATCGCCCGACGACACGACGACGTGGGCTCTCG 120
QY 41 GlyTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
DB 121 GAGTACCAACACACACTCCGCGCACTAGCTCCCACTGCGCGGCTCCATCATCCAG 180
QY 61 ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyValAsnGluArgSerGln 80
DB 181 CCCAGCGCGGAGGGCCCTCCCTGCTGTGAGTTCCAGCCCGGAAATGAACGCTCCAG 240
QY 81 GlnLeuHisLeuArgProGlySerHisSerTyrLeuProGlyIleuGlyIleuSerGlnMet 100
DB 241 GAGCTCCACTCGCGGCAAGTCCCACTACTACCTCCGAGCTGGGGAAGTACAGATG 300
QY 101 GlnPheIleGlySerIleArgProArgLeuGlnLeuLeuProAspProLeuLeuArgPro 120
DB 301 GAGTTCATTGAAAGCAAGCGCCCTCGGCTAGAGCTGCTGCTGACCCCTGCTGCAGCG 360
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrIleAspArgSer 140
DB 361 TCACCCCTCTCGGCCAGCGGCCAGCTCGGGATTCGAAGACTTCACCAAGGACCGTACG 420
QY 141 LeuThrGlyIleuLeuGlnProValSerProProSerProProHisThrAspProGlyLeu 160
DB 421 CTGACGGGGAGAGTGAACCGGTGTCTCCCGCCAGCCCGCCGACACTGACCTTGAGCTG 480
QY 161 GlnLeuValProProArgLeuSerIleGlnGlnLeuIleGlnAsnMetAspArgValAsp 180
DB 481 GAGCTGTGCGCCGACGCGCTGTCCAGAGAGAGACTGATCCAGAACTGAGCCGCGTGAAC 540
QY 181 ArgGlnIleThrMetValGlnGlnGlnIleSerIleLeuLeuIleValGlnGlnGlnLeu 200
DB 541 CAGAGATACCAATGATAGAGCAGAGATCTTAACTGAAGAAAGAGCAGCAACGCTG 600
QY 201 GlnGlnGlnAlaAlaIleAspProProGluProGluIleProValSerProProIleGln 220
DB 601 GAGGAGAGGCTGCCAAGCGCCGAGCTGAGAAAGCCGTGTACCGCCCATCGAG 660

QY 221 SerIleHisArgSerLeuValGlnIleIleTyrAspGluAsnArgIleValAla 240
DB 661 TCGAAGCACCGCAGCTGTGTGACATCTACGAGAACCGGAAGAGCTGAAGCT 720
QY 241 AlaHisArgIleLeuGlnIleuGlyProGlnValAlaLeuProLeuTyrAsnGlnPro 260
DB 721 GCACATCGAATTCGGAGAGCCTTGGGGCCCGAGGTGAGCTGCCGTGTACACAGGCC 780
QY 261 SerAspThrArgGlnTyrHisGluAsnIleIleAsnGlnAlaMetArgIleValLeu 280
DB 781 TCCGACACCCGGCAGATCATGAGAAACATCAAAATTAACAGCGCATGCGGAAGACTA 840
QY 281 IleLeuTyrPheIleArgArgAsnHisAlaArgIleGlnIleProGlyIlePheCysGln 300
DB 841 ATCTGTACTTCAAGAGGAGGATCACTCGGAAACAATGAGAGCAAGTTCTGCCAG 900
QY 301 ArgTyrAspGlnLeuMetGlnAlaLeuGlnIleValIleGlnAsnAspPro 320
DB 901 CCTATGACCACTCATGAGGCTTGGAAGAAAAAGTGAAGCATCGAAGAAACAGCCG 960
QY 321 ArgArgArgAlaIleGlySerIleValArgGlyTyrGlyIleGlnPheProGlyIle 340
DB 961 CCGCGCGCGGCCAAGAGAGCAAGGTGCGGAGTACTACGAAAGCAGTTCCCTGAGATC 1020
QY 341 ArgIleGlnArgIleuGlnIleuArgMetGlnSerArgValGlyGlnArgGlySerGly 360
DB 1021 CCGAAGCAGCGGAGCTGACAGAGCGCATGACAGAGGAGGTGGCCAGCGGCGAGTGG 1080
QY 361 LeuSerMetSerAlaAlaArgSerGlnHisGlnValSerGlnIleLeuAspGlyLeuSer 380
DB 1081 CTGTCTCATCTCGGCGCCCGCAGCAGCAGAGTGTACAGATCATTCATGAGCTCTCA 1140
QY 381 GlnGlnGlnAsnLeuGlnIleuGlnMetArgGlnLeuAlaValIleProMetLeuTyr 400
DB 1141 GAGCAGAGAACCTGGAGAAAGCAGATGCGCAGCTGCGCGTATCCCGCCATGCTGTAC 1200
QY 401 AspAlaAspGlnGlnArgIleIleYsPheIleAsnMetAsnGlyLeuMetAlaAspPheMet 420
DB 1201 GACGCTGACACACAGCGCATCAAGTTCACTCAATGAAGGGCTTATGGCCGACCCATG 1260
QY 421 LysValIleTyrLysAspArgGlnValMetAsnMetIlePheGlnGlnIleuGlySerIlePhe 440
DB 1261 AAGGTGTACAAAGACCGCAGGTCTATGAACATGTGAGTACAGACAGAAAGAGACTTC 1320
QY 441 ArgGlnIlePheMetGlnHisProIleAsnPheGlyLeuIleAlaSerPheLeuGlnArg 460
DB 1321 CCGAGAAAGTTATGACAGATCCCAAGAACTTGGCTGATCGCATCTCCCTGAGAGG 1380
QY 461 LysThrValAlaGlnCysValLeuTyrTyrTyrLeuThrIleLysAsnGluAsnIleLys 480
DB 1381 AAGACAGTGTGAGTGTGCTCTTATTACTACCTGACTTAAGAAATATAGAACTTAAG 1440
QY 481 SerLeuValArgArgSerTyrArgArgArgGlyIleSerGlnGlnGlnGlnGln 500
DB 1441 AGCTGTGTAGACGAGAGCTATGTGGCCCGCGCAGAGCCAGCAGCAGAAACAGCAGG 1500
QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
DB 1501 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1560
QY 521 AspGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 540
DB 1561 GATGAGAAAGAGAAAGAAAGAGGCGGAGAAAGAGAGAGAAAGCCGAGGTGGAGAAC 1620
QY 541 AspLysGluAspLeuLeuIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 560
DB 1621 GACAGAGAAAGACTCTCTCAAGAGAAAGACAGACGACACTTCAGGGAGAGAAAGAG 1680
QY 561 LysGlnAlaValAlaSerIleGlyArgIleThrAlaAsnSerGlnIleArgArgIleGly 580
DB 1681 AAGAGGCTGTGGCTCCAAAGCGCGCAAACTGCCAAACAGCCAGGAGAGAGCAGCAAGGC 1740
QY 581 ArgIleThrArgSerMetAlaAsnGlnIleAsnSerGlnIleAlaIleThrProGlnGln 600

```
Db 1741 CGCATCAACCCGCTCAATGCTAATAGGCCCAACAGCGGAGGCCATCAACCCCCAGAG 1800
Qy 601 SerIaGIuEuAlaSerMetGIuLeuAnGIuSerAgtTPTThGIuGIuMet 620
Db 1801 AGCCCGAGGCTGGCTCCATGAGCTGATAGATTCTCGCTGACGAAGAAGAAATG 1860
Qy 621 GIuThraIaIyBGIyLeuLeuGIuHIaGIyAgtAntPSeRaIaIaIaArgMet 640
Db 1861 GAACACGCAAGAAAGGTCTCTTGAACAACGCCGCAACTGTGCGCATGCCCCGAGATG 1920
Qy 641 ValGIySerIyThraIaIySerGIuCyAlyBaAnPheTyRPhaAnTyRlyLyAaArgGIu 660
Db 1921 GTGGGCTCCAAAGACTGTGTCCGACTGTAACTTCACTTCACTCACTCAAAAGAGGAG 1980
Qy 661 AsnLeuAnPGIuIleuGIuGIuHIaIyBGIySerMetGIuYyBGIuAgtAAnIaArg 680
Db 1981 AACCTCGATGAGATCTTGACAGACACAAAGCTGAAGATGAGAGAGAGAGAGAGCGCGG 2040
Qy 681 ArgIyBGIyBGIyBGIyAIProAlaIaIaIaSerGIuGIuAaIaIaPheProProValIa 700
Db 2041 AGGAAAGAAAGAAAGAGCGCGCGCGCGCGCGCGAGAGAGCTGCAATCCCGCTGGTG 2100
Qy 701 GIuAnPGIuGIuMetGIuAaIaSerGIyValSerGIyAAnGIuGIuMetValGIuGIu 720
Db 2101 GAGGATGAGAGATGAGAGCGCTCGGCGTGAGCGGAATGAGAGAGATGCTGAGAGAG 2160
Qy 721 AlaGIuAlaIeUHIaIaSerGIyAAnGIuValIaProAnGIyGIuCyBSeGIyProAla 740
Db 2161 GGTGAACCTTAACTGCTCTGGAATAGAGGTGCCAGAGAGGAGATCAAGTGGCCAGCC 2220
Qy 741 ThrValaAnAnSerSerAepThrGIuSerIleProSerProHISThrGIuAaIaIy 760
Db 2221 ACTGTCAACAACAGCTACAGACCGAGAGCATCCCTCTCTCTCACTGAGGCGCGCAAG 2280
Qy 761 AspThrGIyGIuAnGIyProLySProProAlaThraIeUGIyAaPGIyProProPro 780
Db 2281 GACACAGGAGCAATGGGCCCAAGCCCGCCAGCCACTGGGCGCGAGCGGCCACCCCA 2340
Qy 781 GIyProProThrProProAgtAgtThraIaIaIaIaIaIaIaIaIaIaIaIaIaIa 800
Db 2341 GGGCCACCCACCCACCAAGAGAGATCCCGGGCCCGCCATGAGCCACCCCGGCTCT 2400
Qy 801 GIuAlaThraGIyAIProThrProProProAlaProProSerProSeRaIaIaIaIa 820
Db 2401 GAAGCCACCGGAGGCCCTTACGCCCCCAACAGACCCCATGCTGCACTCTCTCT 2460
Qy 821 ValIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 840
Db 2461 GTGGTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2520
Qy 841 GIuGIuGIuIyBProProAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 860
Db 2521 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2580
Qy 861 ProValIySerGIuCyThraIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 880
Db 2581 CCGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2640
Qy 881 GIuAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 900
Db 2641 GAGGCGCGCTGAGGCAAGCGCGAGGCGCGCTCAAGCAGAGAGAGAGAGAGAGAG 2700
Qy 901 GIyAaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 920
Db 2701 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2760
Qy 921 CySeSerIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 940
Db 2761 TGCAGTGCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2820
Qy 941 ArgProSerIeUThraIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 960
Db 2821 AGGCCAGCCTCTCAACCCGAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2880
Qy 961 LeuAnPLeuYBGIuIeUyBGIuAgtAaIaIaIaIaIaIaIaIaIaIaIaIaIa 2900
Db 2881 CTGAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2940
Qy 981 ValHIaGIuBProProAgtGIuAaPAlaIaIaIaIaIaIaIaIaIaIaIaIaIa 1000
Db 2941 GTCCATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3000
Qy 1001 ProProGIuAnIeUAnIeUAnIeUAnIeUAnIeUAnIeUAnIeUAnIeUAn 1020
Db 3001 CCACCGCAAAACCTGCAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3060
Qy 1021 GIyBSeSerIySerProAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 1040
Db 3061 GGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3120
Qy 1041 LySLeuProGIyAaPProProCySTPThraIaIaIaIaIaIaIaIaIaIaIaIa 1060
Db 3121 AAGCTGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3180
Qy 1061 GIuValIaIeUyAaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 1080
Db 3181 GAGGTATCAAGGCTCTCCGCAATGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240
Qy 1081 GIyHIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 1100
Db 3241 GGTACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3300
Qy 1101 ThrIleSeRaAnPProProProLeuIleSeSeRaIaIyHIaIaIaIaIaIaIa 1120
Db 3301 ACCATCTCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3360
Qy 1121 GIuIleGIyAaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 1140
Db 3361 CAATAGTGTCAATCTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3420
Qy 1141 AlaIyBAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 1160
Db 3421 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3480
Qy 1161 LeuAlaProPheSeGIyValIyBGIuGIuIleUSeSerProAgtGIyAaIaIa 1180
Db 3481 CTGAGACCTCTTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3540
Qy 1181 ProGIuSerIeUGIyAaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 1200
Db 3541 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600
Qy 1201 GIySerValIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 1220
Db 3601 GGTCTAGTTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660
Qy 1221 SerAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 1240
Db 3661 AGCGCATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3720
Qy 1241 GIyThraIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 1260
Db 3721 GGCACCATCAACAGAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3780
Qy 1261 AspSerIeUProIyBGIyHIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 1280
Db 3781 GACAGCTGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3840
Qy 1281 GIuGIyMeSeSerValIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 1300
Db 3841 GAGGAGTGCATGTCTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3900
Qy 1301 ProHIaGIuThraIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 1320
Db 3901 CCCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3960
```

QY 1321 AlaIleSerSerAlaSerIleGluGlyLeuMetGlyValGluAlaIleProProGluValArgHis 1340
DB 3961 GCCATCTCCTCAAGCCAGCATCGAAGGCTCATGAGCGGTGCATCCCGCGGAGCGACAC 4020
QY 1341 SerProHisHisIleuIleuGluGluGlnHisIleIleArgGlySerIleThrGlnGlyIlePro 1360
DB 4021 AGCCCCACACCTCAAGAGAGACACCATCCGGGGTCCATCAACACAGGGAGATCCCT 4080
QY 1361 ArgSerTyValGluIleGlnGluAspTyLeuValArgGluAlaIleuSerLeuIleuArg 1380
DB 4081 CGGTCTTAGTGAGGACACAGAGACATCTCCGTCCGGAGGCCAAAGCTCTCAAGCCG 4140
QY 1381 GluGlyThrProProProProProSerArgAspLeuThrGluAlaTyIleThrGln 1400
DB 4141 GAGGGACCCCTCCGCCCAACCGCCCTCAACGGGACCTGACCGAGGCTTACAAAGACGAC 4200
QY 1401 AlaIleuGlyProleuIleuSerProAlaHisGluGlyLeuValAlaThrValIleuGlu 1420
DB 4201 GCCCTGGGGCCCCGTGAAGCTGAAGCCGGCCCATGAAGGGCTGGTGGCCACGGTGAAGAG 4260
QY 1421 AlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuPro 1440
DB 4261 GCGGGCGCTCATCATAGATCCCGCGGAGAGCTCGGACACGCCGAGCTGCC 4320
QY 1441 LeuAlaProArgProleuIleuSerGluSerIleThrGlnGlyThrProleuIleuTyAsp 1460
DB 4321 CTGGCCCCGGCGGCTCAAGAGAGGCTCATACACGAGGACCCCGCTCAAGTACGAC 4380
QY 1461 ThrGlyAlaSerThrThrGlySerIleuIleuHisAspValArgSerIleuIleGlySerPro 1480
DB 4381 ACCGGGGCGTCCACCACTGGCTCCAAAGACAGAGTACGCTCCCTCATCGGACACCCC 4440
QY 1481 GlyArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGlu 1500
DB 4441 GCGCGGACGTTCCACCCGTGCACCCGCTGGATGTATGACCCACCGCCCGGACCTGGAA 4500
QY 1501 ArgAlaCysTyrgIleuIleuSerIleuIleuSerArgProGlyThrAlaSerSerSerGly 1520
DB 4501 CGTGCTGTGCTACGAGAGAGCTGAAGAGCCGCGCACGCGCACGCGCTGGGGGGG 4560
QY 1521 SerIleAlaArgGlyAlaProValIleValProGluLeuGlyIleuProArgGlnSerPro 1540
DB 4561 TCCATTGCGCGCGCGCCCGGTCAATTGTGCTGACCTGGTAAAGCCGCGCGACACCCC 4620
QY 1541 LeuThrTyrgIleuAspHisGlyAlaProPheAlaGlyHisIleuProArgGlySerProVal 1560
DB 4621 CTGACCTTAGAGACACAGGGGACCTTTGCGGACCTCCACAGAGTTGCGCGTG 4680
QY 1561 ThrMetArgGluProThrProArgLeuGlnGlySerLeuSerSerSerIleuAlaSer 1580
DB 4681 ACCATCGGGAGGCCACGCGCGCTGCAGAGGAGGACCTTTGCTGCAGCAAGCATCC 4740
QY 1581 GlnAspArgGlyLeuThrSerThrProArgGluIleAlaIleuSerProHisSerThrVal 1600
DB 4741 CAGGACCGGAAGCTGACGTGACGCTCTGTGATGACCGCAAGTCCCGGACACACACCGTG 4800
QY 1601 ProGlnHisHisProHisProIleSerProTyrgIleHisIleuIleuArgGlyValIleuGly 1620
DB 4801 CCGGACACACACACACACACCTGACCTGATGAGACCTGCTCGGGCGGAGGGG 4860
QY 1621 ValAspLeuTyArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGly 1640
DB 4861 GTGGAGCTGTATGAGGACACATCCCGCTGGCTTGGACCCCACTCCATCAACCCCGGG 4920
QY 1641 IleProLeuAspAlaAlaAlaAlaTyTyTyLeuProArgHisIleuAlaProAspProThr 1660
DB 4921 ATCCCTCTGAGACACACCGCTGCTTACTGCTCCCGGACACCTGGCGCCCAACCCAC 4980
QY 1661 TyTyProHisIleuTyTyProProTyTyLeuIleAlaTyTyTyProAspThrAlaAlaLeuGlu 1680
DB 4981 TACCCGACCTGTACCAACCTACTATCCGGGCTACCCGACACAGGGGCGCTGGAG 5040

QY 1681 AsnArgGlnThrIleIleHisAspTyTyIleThrSerGlnGluMetHisHisAsnThrAla 1700
DB 5041 AACCGCACACCATCATCATATATCATCATCTGCACCAATGTGACCAACAGGCGC 5100
QY 1701 ThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeu 1720
DB 5101 ACCGCGATGGCCACGAGACTATATGCTGAGGGGCTTCGCGCCCGGACAGTCTCGCTG 5160
QY 1721 AlaLeuAsnTyAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeu 1740
DB 5161 GCACCTCACTACGCTCGGGTCCCGGAGCATCATGACCTGTGCCAAGTGCACACCTG 5220
QY 1741 ProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTy 1760
DB 5221 CCGTGCTGTGCCCCCGCACACAGGACCCCAAGCCACCATGATGACCGCTTGTCTAC 5280
QY 1761 LeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProleuSerProGly 1780
DB 5281 CTCCCAACCGCGCCCGAGCCCTTCAGACACCCGACAGACCTCCCACTTCCACAGA 5340
QY 1781 GlyProThrHisIleuThrTySerProThrThrThrSerSerSerGluArgIleArgAspArg 1800
DB 5341 GGTCCACACACTGTACAAACCAACCAACGATCCTCGTCCGAGGGGAGCGACACCG 5400
QY 1801 AspArgGluArgAspArgAspArgGluArgGluIleuSerIleIleuThrSerThrThrThr 1820
DB 5401 GATCGAGACCGGACCGGAGTGGGAGCGGAAATCATCTCACTCCATCCACAGACG 5460
QY 1821 ValGlnHisAlaProIleThrArgProGlyTyThrGluGlnSerSerGlySerSerGlySer 1840
DB 5461 GTGAGACAGCACCCATCTGAGACCTGTGACAGCTGTACAGACAGACAGCGGACGCGCAC 5520
QY 1841 SerGlyGlyGlyGlyGlySerSerSerSerArgProAlaSerHisSerHisAlaHisGlnHis 1860
DB 5521 AGCGGGGGGGGGGGGAGAGAGCGCGCCGCTCCCATCCATGCCACCAACGAC 5580
QY 1861 SerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsn 1880
DB 5581 TCGCCCATCTCCCTCGAGCCAGAGTGCCTCCAGACAGACCCCATGTGTCTTCAAC 5640
QY 1881 ThrGlyMetIleGlyIleIleThrAlaValGluProSerTyProThrValLeuArgSer 1900
DB 5641 ACHGGCATAGAGGATATCATACCGTGTGAGGCCAGAGACCCAGCGTCTGAGTTC 5700
QY 1901 ThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCysPro 1920
DB 5701 ACTTCACCTCTCATACCGCTTGCAGCTGCACATTCACCACTGCGACCACTGCCA 5760
QY 1921 LeuGlyGlyThrLeuAspGlyValTyTyProThrLeuMetGluProValLeuLeuProIle 1940
DB 5761 CTGGGGGACACCTCGATGGGGTCTTACCTTACCTCATGAGAGCCGTCTTGTGCGCAAG 5820
QY 1941 GlnAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLeu 1960
DB 5821 GAGGCCCCCGGGGTGCGCGGACAGCGGCCCCGAGACACGCGCATGCTTCTTC 5880
QY 1961 AlaTyProProAlaArgSerGlyLeuGluProAlaSerSerProSerTyIleGlySerGln 1980
DB 5881 GCCAAGCCCCCAGCCCGCTCCGGGCTGAGCCGCGCTCTCCCCAGCAAGGCTCGAG 5940
QY 1981 ProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaIle 2000
DB 5941 CCGGGCCCTTAGTGCTCTGTCTGTGCGACAGCCACATGCGCCGACCCCTGGGAG 6000
QY 2001 AsnLeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSerAsp 2020
DB 6001 AACCTCGACCTTCACACCGCAGCCGAGCCGCGGCGCACCTGCGCTGAGAC 6060
QY 2021 ProHisArgGluTyThrGlnSerTyTyProPheSerIleGlnGluLeuGluIleuArgSer 2040
DB 6061 CCGCACCGGAAAGACTCAAGTAAACCTTTTTCATTCAGGAATGGAATCCGCTTCT 6120
QY 2041 LeuGlyTyTyHisGlySerSerTySerProGluGlyValGluProValSerProValSer 2060

Db 6121 CTGGGTTACCAACGACAGCTACAGCCCGAAGGGGTGAGCCCGTACAGCCCTGTAGGC 6180
 Qy 2061 SerProSerLeuThrHisAspArgGlyLeuProGlyHisGlyLeuGlyGlyLeuAspGlySer 2080
 Db 6181 TCACCCAGCTGACCCACGACAAAGGGGTCCCAAGACCTGGAAAGAGCTGCACAAAGGC 6240
 Qy 2081 HisLeuGlyGlyLeuArgProGlyGlnProGlyProValIlybLeuGlyGlyGlyGly 2100
 Db 6241 CACCTGAGGGGAGAGTGGGGCCCAAGAGCCAGCCCGGTGAAGCTTGGGGGAGGCC 6300
 Qy 2101 AlaHisLeuProHisGlyLeuArgProLeuProGlyLeuSerGlnProSerSerSerProLeu 2120
 Db 6301 GCCCACCCTCCACACCTGGGGCCGCTGCTGAGAGCAGCCCTGCTGACGCCCTCTCTC 6360
 Qy 2121 GlnThrAlaProGlyValIlybGlyHisGlnArgValIlybLeuAlaGlnHisIleSer 2140
 Db 6361 CAGACCCGCCCAAGGGGTCAAAAGGTACACGGGGTGTGCACCTGGGCCAGCATACAGT 6420
 Qy 2141 GlnValIleThrGlnAspGlyThrArgHisHisProGlnGlnLeuSerAlaProLeuPro 2160
 Db 6421 GAGGTATCACACAGACTACACCCGGCACACCCACAGACTCAGCGACCCCTGCCC 6480
 Qy 2161 AlaProLeuTySerPheProGlyAlaSerCybProValLeuAspLeuArgArgProPro 2180
 Db 6481 GCCCCTCTACTCTCTCCCTGGGGCCAGCTGCCCTGCTGACCTCCGCCGCCACCC 6540
 Qy 2181 SerAspLeuTyLeuProProProProAspHisGlyValAlaProAlaArgGlySerProHisSer 2200
 Db 6541 AGTGACCTTACCTCCGCCGCCGACCATGTGCCCCGCCCTGCTCTCCCAAGC 6600
 Qy 2201 GlnGlyGlybAspSerProGlnProAsnIlybSerValLeuGlyGlyGlyGlyAsp 2220
 Db 6601 GAAGGGGGCAAGAGTCTCCAGAGCCAAACAGCGCTGGTCTTGGGTGGTGGAGAC 6660
 Qy 2221 GlyIleGlnProValSerProProGlyGlyMetThrGlnProGlyHisIleSerArgSerAla 2240
 Db 6661 GGTATTGAACCTGTGCTCCACCGAGGAGTGAACGAGCGAGGACTCCCGGAGTGCT 6720
 Qy 2241 ValIlybProLeuLeuTyArgAspGlyGlnGlnThrGlnProSerArgMetGlySerIlyb 2260
 Db 6721 GTGTACCCGTGCTGTCAGCGAGATGGGAAACAGCGAGGCCAGCAAGATGGCTCCAG 6780
 Qy 2261 SerProGlyAsnThrSerGlnProProAlaPhePheSerIlybLeuThrGlnSerAsnSer 2280
 Db 6781 TCTCCAGGCAACACGAGCCGCGCCCTCTTCCAGCAAGCTGACCGAGAGCAATCC 6840
 Qy 2281 AlMetValIlybSerIlybGlnGlnIleAsnIlybLeuAsnThrHisAsnArgAsn 2300
 Db 6841 GCCATGGTCAAGTCCAAAGCAAGAGATCAACAAAGAGCTGAACACCCACAAACCGGAT 6900
 Qy 2301 GlnProGlnTyArgHisIleSerGlnProGlyThrGlnIlePheAsnMetProAlaIleThr 2320
 Db 6901 GAGCCTGAATACATATACAGCCGAGGAGAGTCTTCAATATAGCCCGCCATCACC 6960
 Qy 2321 GlyThrGlyLeuMetThrTyArgSerGlnAlaValGlnGlnHisIleSerThrAsnMet 2340
 Db 6961 GGAACAGGCTTATGACCTTATAGAGCCAGCGGTGAGAGAACATGCCACACCAACATG 7020
 Qy 2341 GlyLeuGlnAlaIleIleArgIlybAlaLeuMetGlyIlybTyArgGlnIlybGlnGlnSer 2360
 Db 7021 GGCTGAGAGGCATTAATTAGAAAGGCACTCATGGTAAATATAGCAAGTGGAAAGTCC 7080
 Qy 2361 ProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAla 2380
 Db 7081 CCGCCGCTCAGCGCAATGCTTTAACTCTGATATCCAGTCCAGACCTGCCCGCTGCT 7140
 Qy 2381 MetProIleThrAlaAlaAspGlyArgSerAspHisHisLeuThrSerProGlyGlyGly 2400
 Db 7141 ATGCCCATTAACCGCTGCTGACGAGAGTGAACACACATCTCAGCTGCCAGGTGGGGC 7200
 Qy 2401 GlyIlybAlaIlybValSerGlyArgProSerSerArgIlybAlaIlybSerProAlaProGly 2420

Db 7201 GGGAGGCCAAGGCTCTGACAGACCCAGACGCCGAAAGCAAGTCCCGGCCCGGCGC 7260
 Qy 2421 LeuAlaSerGlyAspArgProProSerValIleSerSerValHisSerGlyGlyIlybAspCybAsn 2440
 Db 7261 CTGGCATCTGGAGACCGGCCACCTCTGTCTCTCAATGACCTCGAGAGAGACTGCAAC 7320
 Qy 2441 ArgArgThrProLeuThrAsnArgValIlybGlnAspArgProSerSerAlaGlySerThr 2460
 Db 7321 CGCGGAGCGCGCTCACCAACCGGTGTGGAGAGACAGGCCCTGCTCGGAGGTTCCACG 7380
 Qy 2461 ProPheProTyArgProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProPro 2480
 Db 7381 CCAITCCCCCTAACACCCCTGATCATGCGCTGACGGCGGTGTCTATGCTTCCCAACC 7440
 Qy 2481 ProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisIleATPAspGly 2500
 Db 7441 CACCGGGCTCTCCCGGGGAGGGGCGCTGCTGGGCCCAACACGCGCTGGAGCAG 7500
 Qy 2501 GlnProIlybProLeuLeuCybSerGlnTyArgIlybThrLeuSerAspSerGln 2517
 Db 7501 GAGCCCAAGCACTGCTCTGCTGCAATACGACACTCTCCAGACGAG 7551
 Db
 RESULT 2
 ADG86298 standard; cDNA; 8561 BP.
 ID
 XX
 AC ADG86298;
 DT 11-MAR-2004 (first entry)
 XX
 XX
 DE Human SMRT encoding cDNA SEQ ID NO:12.
 XX
 KW SMRT; silencing mediator for retinoid and thyroid hormone action;
 KW SMRT inhibitor; cytosolic; antiinflammatory; antiarthritic;
 KW antirheumatic; antisense therapy; inflammatory disorder;
 KW rheumatoid arthritis; hyperproliferative disorder; cancer; leukaemia;
 KW breast cancer; human; gene; ss.
 XX
 OS Homo sapiens.
 XX
 FH
 FT Key
 FT CDS
 FT /*tag= a
 FT /product= "SMRT"
 XX
 PN NO2003106645-A2.
 XX
 PD 24-DEC-2003.
 XX
 PF 17-JUN-2003; 2003MO-US018923.
 XX
 PR 17-JUN-2002; 2002US-00174014.
 XX
 PA (ISIS-) ISIS PHARM INC.
 XX
 PI Bennett CF, Freier SM, Doble KM;
 XX
 DR WPI; 2004-082184/08.
 DR P-PDB; ADG86299.
 DR GENBANK; NM_006312.
 XX
 PT Novel antisense compound targeted to nucleic acid encoding SMRT
 PT (silencing mediator for retinoid and thyroid hormone action), useful for
 PT treating animal having disease associated with SMRT such as cancer,
 PT rheumatoid arthritis.
 XX
 PS Example 15; SEQ ID NO 12; 260bp; English.
 XX
 CC The present invention describes a compound (I) 8-50 nucleobases in length
 CC targeted to a nucleic acid molecule encoding SMRT (silencing mediator for
 CC retinoid and thyroid hormone action), where (I) specifically hybridises
 CC with the nucleic acid molecule encoding SMRT and inhibits expression of a
 CC SMRT. (I) specifically hybridises with at least 8-nucleobase portion of a

CC preferred target region on nucleic acid molecule encoding SMRT. Also
CC described is a composition (II) comprising (I) and a carrier or diluent.
CC (I) and (II) have cytostatic, antiinflammatory, antiarthritic and
CC antirheumatic activities, and can be used in antisense therapy, and as
CC SMRT expression inhibitors. (I) is useful for inhibiting the expression
CC of SMRT in cells or tissues. (I) is also useful for treating an animal
CC having a disease or condition associated with SMRT, e.g., inflammatory
CC disorder such as rheumatoid arthritis, or a hyperproliferative disorder
CC such as cancer chosen from leukaemia and breast cancer, by inhibiting the
CC expression of SMRT. (I) is useful for diagnostics, therapeutics,
CC prophylaxis and as research reagents and kits. The present sequence
CC encodes human SMRT, which is used in an example from the present
CC invention. N.B. The present sequence is designated as SEQ ID NO:11 in
CC example 15 but corresponds to SEQ ID NO:12 in the Sequence Listing.

SO Sequence 8561 BP; 1862 A; 3033 C; 2525 G; 1141 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0	Length:	8561
Score:	13215.00	Matches:	2517
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	12	Gaps:	0

US-09-522-753-5 (1-2517) x ADG86298 (1-8561)

```
QY 1 MetSerG1SerThGlnLeuValAlaGlnThrProAlaThrGluProArgTyrPro 20
Db 2 AAGTCGAGGCTCCACACAGCTTGAGCAGAGCGAGGACCTAGCCCGCTACCCG 61
QY 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeu 40
Db 62 CCCACAGGCTTTCCTACCCAGTGCAGATCGCCGAGCGACGACCGACCGGCTCTCG 121
QY 41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
Db 122 GAGTACACACACACACCTCCGCGAGCTAGCTCCACCTGCTCGCGGCTCCATCACCAG 181
QY 61 ProGlnArgArgProSerLeuSerGluPheGlnProGluYanGluArgSerGln 80
Db 182 CCCACGCGGCGAGGCGCTCTCTGCTGCTGAGTTCACGCGCGAGATACCGATCCAG 241
QY 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlySerSerGluMet 100
Db 242 GAGCTCCACCTGGCGGCGCAGAGTCCCATCTACCTGCCGAGCTGGGAGAGTCAGAGATG 301
QY 101 GluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
Db 302 GAGTTCATATGAAAGCAAGGCGCTCGCTAGAGCTCTGCTGACCCCTGCTGCACCG 361
QY 121 SerProLeuLeuAlaThrGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
Db 362 TCACTCCCTCTGGCCACGCGGCGAGCTCGGAGATCTGAAGACTTACCAAGAGCCATTAGC 421
QY 141 LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
Db 422 CTGACGCGGCAAGCTGGAACCGAGTGTCTCCCGCCAGCCCGCCGACACTGACCTGAGCTG 481
QY 161 GluLeuValIleProProArgLeuSerLysGluGluLeuIleGlnAsnMetLeuAspValAsp 180
Db 482 GAGCTGGTCCCGCAGCGGCTGTCCAAGGAGGAGCTATCCAGAACTGAGCCCGCTGGAGAC 541
QY 181 ArgGluIleLeuThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeu 200
Db 542 CAGAGATATCACTAGTGAAGCAGAGATCTTAAGCTGAAGAGAGAGCAGCAACGACTG 601
QY 201 GluGluGluAlaAlaLysProProGluProGluLysProValSerProProProIleGln 220
Db 602 GAGGAGAGGCTGTCGAAGCGCGCCGAGCTGAGAGAACCGCTGTCACTGCGCGCCATCGAG 661
QY 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysValAlaGluAla 240
```

```
Db 662 TCGAAGCACCGCAGCTGGTGCAATCATACAGCAGAGAACCGGAGAAAGCTGAAGCT 721
QY 241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrHisGlnPro 260
Db 722 GCACATCCGATTCTGGAAAGGCTGGGGGCCAGAGTGGAGCTCCGCTGTACAAACGAGCC 781
QY 261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
Db 782 TCCGACACCCCGCAGTATCATATAGAAACATCAAAATMAACAGAGCATGGGAGAAAGCTA 841
QY 281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGln 300
Db 842 ATCTTGTACTTAAGAGAGAGATACGCTCGAGAAACAATGAGAGACAGAAATTTCTCCAG 901
QY 301 ArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnPro 320
Db 902 CGCTATGACACAGCTCATGAGAGGCTTTGGAAAAAGGTGAGACGCTGCAAAAACACCCG 961
QY 321 ArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340
Db 962 CGCCGCGCGGCGCAAGAGAGCAGAGTGGCGGAGTACTACGAAAAAGCATTTCCCTGAGATC 1021
QY 341 ArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyIleArgLysSerGly 360
Db 1022 CGCAAGCAGCGGAGCTGAGAGAGCGCATGCAAGAGAGGCTGGCCAGCGGCGAGTGGG 1081
QY 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
Db 1082 CTGTCTCATGTGGCGCGCGCGGAGCGACGAGCGAGGTGTGAGATCATGATGAGCTCTCA 1141
QY 381 GluGlnGluAsnLeuGlnLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400
Db 1142 GAGCAGAGAGAACCTGAGAGAGCAGATGCGCGAGCTGGCGGTATCCCGCCATGCTGATC 1201
QY 401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspPromet 420
Db 1202 GAGGCTGACACAGCGCATCAAGTTCATCAATCAATGAAAGGCTTATGGCCACCCCATG 1261
QY 421 LysValTyrLysAspArgGlnValMetAsnMetTyrSerGluGlnGluLysGluThrPhe 440
Db 1262 AAGGTGTCAAAAGACCGCAGGTCATGAACATGTGAGTGAAGCAGAGAAAGAGACTTTC 1321
QY 441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
Db 1322 CGGAGAAAGTTTCATGACAGATCCCAAGAACTTTGGCTGTATGCAATCTCTCGAGAG 1381
QY 461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480
Db 1382 AAGACAGTGGCTGAGTGGCTCTTATTTACTGACTTAAGAAATAGAAACTTAAAG 1441
QY 481 SerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGln 500
Db 1442 AGCTGTGTAGAGGAGAGCTATGCGCGCGCGCAAGAGCAGACAGCAACAGCAGCAG 1501
QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
Db 1502 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1561
QY 521 AspGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLys 540
Db 1562 GATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1621
QY 541 AspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyLysAspAspGlu 560
Db 1622 GACAAAGAAAGACTCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1681
QY 561 LysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGly 580
Db 1682 AAGGAGGCTGTGCTCCCAAAAGCGCAAAATGCGCAACAGCAGGAGAGAGAGCAAAAG 1741
QY 581 ArgIleThrArgSerMetLysAsnGlnAlaAsnSerGluGluAlaIleThrProGlnGln 600
Db 1742 CGCATCAACCGCTCAATGGCTAATGAGCCCAACAGGAGAGGCCATCAACCCCGCAGAG 1801
```

[illegible]

QY	961	LeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLys	980
Db	2882	CTGAACCTCGAAGCAGCTGAGACAGCGGCTCCATCCCGCCCATCCAGTCAACAA	2941
QY	981	ValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaProPro	1000
Db	2942	GTCCATGAGCCCCCGGGAGAGAGCAGACTCCACCAAGCCAGTCCCCCGAGCCACCG	3001
QY	1001	ProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProArg	1020
Db	3002	CCACCGCAAAACCTGCAGCGCGGAGAGCGAGCCCTCGAGCAGCCTGGGACGACGCCCGG	3061
QY	1021	GlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGln	1040
Db	3062	GCGAAGAGCAGAGACCCCGGCACCCCCGCCACAAAGAGGCTTCGACGCGGAGCCGACG	3121
QY	1041	LysLeuProGlyAspProProProCysThrThrSerGlyLeuProPheProValProProArg	1060
Db	3122	AAGCTGCTGGGGAGCCCCCTTGCTGGACTTCGGGCTGGCCCTTCGCCGATGCCCCCGT	3181
QY	1061	GluValAlaIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaProPro	1080
Db	3182	GAGGTGATCAAGGCTCCCGGCATGCCCGGACCCCTCAGCTTCTCTAGCTCAACCT	3241
QY	1081	GlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProPro	1100
Db	3242	GGTCACCCCACTGCCCCCTGGGCTCCATGACACTCCCGGCCGCTTCGGCGCGCCACCC	3301
QY	1101	ThrIleSerAnProProProProLeuIleSerSerAlaLysHisProSerValLeuGluArg	1120
Db	3302	ACCATCTCAACCCGCTCCCTCATCTCTCTGCGCAAGCACCCACAGCTCTGAAAGG	3361
QY	1121	GlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGluHis	1140
Db	3362	CAATTAAGTGGCATCTCCCAAGAAATGCTCGTCAAGCTCAACGTCCTGACTCAGACAT	3421
QY	1141	AlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLys	1160
Db	3422	GCCAAAGCCCCGGGGGCGCTGTCAACATGGGGCTGCCCTGACCATGAGACCCCAAAAG	3481
QY	1161	LeuAlaProPheSerGlyValLysGlnGlnLeuSerProArgGlyGlnAlaGlyPro	1180
Db	3482	CTGGCACCTTCACGCGAGTGAACAGACAGCATGTCCTCCACGGGGCCAGGCTGGGCCA	3541
QY	1181	ProGluSerLeuGlyValProThrAlaGlnGluLysSerValLeuArgGlyThrAlaLeu	1200
Db	3542	CCGAGAGGCTGGGGGGTGGCCACAGCCACAGAGGCGTCCGTGTAAGAGGACAGCTCTG	3601
QY	1201	GlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAsp	1220
Db	3602	GGCTCAGTCCGGGCGGAGAGATCATCCAAAGGATTCACAGACAGGGTGCCTCGGAC	3661
QY	1221	SerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLys	1240
Db	3662	AGCCCATCATCATCCCGCTCATCACCCACGCGACGCGACGTGACGTCTGTAAAG	3721
QY	1241	GlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGlu	1260
Db	3722	GGCACCATCAACAGATCATCGGGAGAGACGCCCGATCGCTTGGACCGCGCGGGAG	3781
QY	1261	AspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyr	1280
Db	3782	GACAGCTGCGCAAGGCGCAAGTATCTTAAGAGGCAAAAGGCGCAGTCTTGCTCAT	3841
QY	1281	GluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyPro	1300
Db	3842	GAGGTGTCATGCTGTGACCCAGTGTCTCAAGAGAGACGCGAAGACACTCAGGACC	3901
QY	1301	ProHisGlnThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValAlaArg	1320
Db	3902	CCCATATGAGCGCGCGCCCAAGCGCACTTAAGACATATGAGAGGGCGCGTGGGACGA	3961
QY	1321	AlaIleSerSerAlaSerIleGlyGlyLeuMetGlyArgAlaIleProProGluArgHis	1340

|||||
Db 3962 GCCATCTCTCAAGCAGCATCAAGAGTCTCATGGGCGTGCATCCCGCGGACGACAC 4021
QY 1341 SerProHisHisLeuLeuLeuGlnGlnHisHisLeuArgLysSerLeuHisGlnGlnLeuPro 1360
Db 4022 ACCCCCCACACCTCTCAAGAGCAGCACAATCCGGGGGTCATCAACACAGGAGATCCCT 4081
QY 1361 ArgSerTyrValGlnValGlnGlnLeuAspTyrLeuArgArgGlnValAlaLeuLeuLysArg 1380
Db 4082 CGGTCTCTAGTGAAGGACACAGAGAGCTACCTGCGTGGGAGGCGCAAGCTCTCAAGCGG 4141
QY 1381 GlnGlyTyrProProProProProProSerArgAspLeuThrGlnAlaTyrLysThrGln 1400
Db 4142 GAGGGACACGCTCCGCCCCACGCGCTCAAGGACCTGACCGAGGCTCAAGAGCGCAG 4201
QY 1401 AlaLeuGlyProLeuLysLeuLysProAlaHisGlnGlyLeuValAlaThrValLysGln 1420
Db 4202 GCCCTGGGGCCCCCTGAAGCTGAAGCGGGCCCATGAAGGCGCTGGTGGCCACGGTGAAGAG 4261
QY 1421 AlaGlyArgSerLeuHisGlnLeuLeuProArgGlnGlnLeuArgHisThrProGlnLeuPro 1440
Db 4262 GGGGGCGGCTCATCATAGATCCCGCGAGGAGCTGGGACACGCGCCGAGCTGCC 4321
QY 1441 LeuAlaProArgProLeuLysGlnGlySerLeuLeuThrGlnGlyTyrProLeuLysTyrAsp 1460
Db 4322 CTGGCCCCCGGCGCTCAAGGAGGGGCTCCATCAACGAGGCGCACCCCGCTCAAGTTCGAC 4381
QY 1461 ThrGlyAlaSerThrThrGlySerLysHisAspValArgSerLeuLeuGlySerPro 1480
Db 4382 ACCGGGCGCTCAACCTGCTCCAAAGACAGACGAGTACGCTCCCTCATCGGACACCCC 4441
QY 1481 GlyArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGln 1500
Db 4442 GCGCGGACGTTCCACCCGTGCACCCGCTGATGTATGGCCGACCCCGGGACCTGGAA 4501
QY 1501 ArgAlaCysTyrGlnGlnSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGly 1520
Db 4502 CGGCTCTGCTCAAGAGAGAGCTGAAGAGCGGCGCAGGACCGCGCAGCAGCTCGGGGCG 4561
QY 1521 SerLeuAlaArgGlyValaProValLeuAlaProGlnLeuGlyLysAspProArgGlnSerPro 1540
Db 4562 TCCATTGCGCGGGCGCGCCCGGCTCATTTGCTGACCTGGGTAAAGCCGGCGCAGAACCCC 4621
QY 1541 LeuThrTyrGlnAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerProVal 1560
Db 4622 CTGACCTATGAGGACCAACGGGACACCTTTGCCGCGCACCTCCACGAGGTTGCCCGTG 4681
QY 1561 ThrMetArgGlnProThrProArgLeuGlnGlySerLeuSerSerSerLysAlaSer 1580
Db 4682 ACCATGCGGAGCCACGCGCGGCTGCAAGAGGAGCTTTGCTCAGCAAGCATCC 4741
QY 1581 GlnAspArgLysLeuThrSerThrProArgGlnLeuAlaLysSerProHisSerThrVal 1600
Db 4742 CAGGACCGAAAGCTGACGTGACGCTCGTGAATGCGCAAGTCCCGGACACACACCGTG 4801
QY 1601 ProGlnHisHisProHisProLysSerProTyrGlnHisLeuLeuArgGlyValaSerGly 1620
Db 4802 CCGGAGACACACCAACCCCATCTGCGCTTATGAGCAGCTGCTTGGGGCGTGAATGGC 4861
QY 1621 ValAspLeuTyrArgSerHisLeuProLeuAlaPheAspProThrSerLeuProArgGly 1640
Db 4862 GTGGACCTGATGAGCGCACATCCCGCTGGCTTGAGACCCGACCTCATACCCCGGCG 4921
QY 1641 IleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAspProThr 1660
Db 4922 ATCCCTCTGGACCGACCGCTGCTTACTGCTGCCGACCCCGACACCTGGCCCCAACCCAC 4981
QY 1661 TyrProHisLeuTyrProProTyrLeuLysArgGlyTyrProAspThrAlaAlaLeuGln 1680
Db 4982 TACCCGACCTGTAACCACTTACTCATCGGGCTACCCCGACGAGGGGGCGCTGGAG 5041
QY 1681 AsnArgGlnThrIleIleLeuAspTyrIleThrSerGlnLeuMetHisHisAsnThrAla 1700

Db 5042 AACGGCAGACCATCATGATGATCATCACTTCGACGAGATGACACCAACAGGCG 5101
QY 1701 ThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGlnSerLeu 1720
Db 5102 ACCGCGATGGCCCAAGGAGCTATATGCTGAGGGGCTCTCGCCCGGACAGTCTCGCTG 5161
QY 1721 AlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeu 1740
Db 5162 GCATCAACTGACGTGCGGGGTCCCGAGGACATCATGACTGTCCCAATGTGCACCTCG 5221
QY 1741 ProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyr 1760
Db 5222 CTGTGCTGTGCCCCGACACCAAGGACCCGACCGCATGGACCGCTTGCTGCTTAC 5281
QY 1761 LeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGly 1780
Db 5282 CTCCCAACGGGCGCCAGCGCTTCAGAGCGGCGCACAGAGCTCCCACTTCCCGAGGA 5341
QY 1781 GlyProThrHisLeuThrLysProThrThrThrSerSerSerGlnArgGlnArgAspArg 1800
Db 5342 GGTCCAAACACCTTGCAAAACCAACCAACACACAGTCTCTCCGAGCGGAGGAGACCG 5401
QY 1801 AspArgGlnArgAspArgAspArgGlnArgGlyLysSerLeuLeuThrSerThrThrThr 1820
Db 5402 GATCGAGAGCGGAGACGGGATCGGAGCGGAAAGTCCATCTTCACTCCACACGACG 5461
QY 1821 ValGlnHisAlaProIleTyrArgProGlyThrGlnGlnSerSerGlySer 1840
Db 5462 GTGAGACAGCACCATCTGAGACCTGGTACAGACAGACGAGCGGACGCGGACG 5521
QY 1841 SerGlyGlyGlyGlyLysSerSerSerArgProAlaSerHisSerHisAlaHisGlnHis 1860
Db 5522 ACCGGGCGGGGTGGGGGCGAGCAGCGCCCGGCTCCCATCTCCATCCACCAACGAC 5581
QY 1861 SerProLysSerProArgThrGlnAspAlaLeuGlnArgProSerValLeuHisAsn 1880
Db 5582 TGCCCATCTCCCTCGACCCAGGATGCTTCCAGACAGAACCCAGTGTTCACAC 5641
QY 1881 ThrGlyMetLysGlyIleIleThrAlaValGlnProSerLysProThrValLeuArgSer 1900
Db 5642 ACGGATAAAGGATTCATCACCGCTGAGGCCAGACGACGACCGGCTCGAGGCTC 5701
QY 1901 ThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCysPro 1920
Db 5702 ACTCCACTCTCTCAACCGCTTGCCAGCTGCACATTCCTCCACCTGCCACCACTGCCA 5761
QY 1921 LeuGlyGlyTyrLeuAspGlyValTyrProThrLeuMetGlnProValLeuLeuProLys 1940
Db 5762 CTGGGGGAGACCTCGATGGGATCTACCTTACCTCATGAGCCGATCTTGCTGCCAAG 5821
QY 1941 GlnAlaProArgValAlaArgProGlnArgProArgAlaAspThrGlyHisAlaPheLeu 1960
Db 5822 GAGGCGCCCGGGTGGCCCGGCGCAGCGGCGCCGAGACACCGGCCATGCTTCTC 5881
QY 1961 AlaLysProProAlaArgSerGlyLeuGlnProAlaSerSerProSerLysGlySerGln 1980
Db 5882 GCCAAGCCCCCAGCGCGCTCGGGCTGAGCGCGCTCTCCCGCGCAAGGGCTCTGGAG 5941
QY 1981 ProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLys 2000
Db 5942 CCGGCGCCCTGATGCTCTGTCTGTGGCAGCGCACCATGCGCCGACCCCTGGGAAG 6001
QY 2001 AsnLeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSerAsp 2020
Db 6002 AACCTCGACCTTACACCGCCAGCCGAGCCGCGCGGCGCCACTGCTCGGCTCGAGC 6061
QY 2021 ProHisArgGlnLysThrGlnSerLysProPheSerLeuGlnGlnLeuGlnLeuArgSer 2040
Db 6062 CCGACCGGGAAAGACTAAAGTAAACCTTTTCATGCAAGAACTGGAACCTCGCTTCT 6121
QY 2041 LeuGlyTyrHisGlySerSerTyrSerProGlnGlyValGlnProValSerProValSer 2060
Db 6122 CTGGGTACACGAGCGACGATTACAGCCCGGAAGGGGTGAAGCCGTCAAGCCCTGTGAGC 6181

QY 2061 SerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSer 2080
 Db 6182 TCACCCAGTCTGACCCAGCAAGAGGGGCTCCCAAGACCTCGAAGAGCTGACAGAGC 6241
 QY 2081 HisLeuGluGluGlyLeuLeuArgProLysGluInProValLysLeuGluGlyGluA 2100
 Db 6242 CACCTGAGGGGGAGCTGCGGCCCAAGCAGCGCCCGGTGAAGCTTGGGGGAGAGCC 6301
 QY 2101 AlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSerProLeu 2120
 Db 6302 GCCCACCCTCCACACCTGCGGCGCTGCTGAGAGCCAGCCCTGCTCAGCCCTGCTC 6361
 QY 2121 GlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisL 2140
 Db 6362 CAGACCGCCCAAGGGGTCGCAAGGTCAACAGCGGTGTCACTCGGCCAGCATCATG 6421
 QY 2141 GluValLleThrGlnAspLysThrArgHisHisProGlnGlnLeuSerAlaProLeu 2160
 Db 6422 GAGGTCAATCACACAGGACTACACCGCCGACCCACAGCAGCTCAGCGCACCTGCCC 6481
 QY 2161 AlaProLeuTyrSerPheProGlyAlaSerCysAspProValLeuAspLeuArgPro 2180
 Db 6482 GCCCCTCTACTCTCTCCCTGGGGCCAGCTGCCCTGCTGACCTCCGCGCCACCC 6541
 QY 2181 SerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerProHis 2200
 Db 6542 AGTACCTTACCTCCCGCCCGCCGACCATGCTCCCGCCGCTGCTCCCGCCACAGC 6601
 QY 2201 GlnGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGluGlyGlyL 2220
 Db 6602 GAAGGGGGCAAGGTCTCCAGAGCCCAACAGACGTGGTCTGGTGGTGAAGAC 6661
 QY 2221 GlnLleGluProAlaSerProProGluGlyMetThrLubProGlyHisSerArgSer 2240
 Db 6662 GGTATTGAACCTGTGTCCACCGAGGAGCATGACGAGCAGGCGCACTCCGAGTGT 6721
 QY 2241 ValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySer 2260
 Db 6722 GTGTACCCGCTGCTGTCGCGGATGGGAAACAGCGAGCCAGCATGGCTCCAG 6781
 QY 2261 SerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsn 2280
 Db 6782 TCTCCAGGCAACCGAGCCGCGCAGCTTCTTCAAGAGCTGACCGAGAGCAATCC 6841
 QY 2281 AlaMetValLysSerLysLysGlnLleAsnLysLysLeuMetThrHisAspArgAsn 2300
 Db 6842 GCCATGGTCAAGTCCAGAGAGCAAGATCAACAGAAAGCTGAACACCCCAACCGAAT 6901
 QY 2301 GluProGluTyrAsnLysSerGlnProGlyThrGluLlePheAsnMetProAlaLle 2320
 Db 6902 GAGCTGAATACATATCAGCCAGCCGCGAGCGAGATCTTCAATATGCCCGCATCAC 6961
 QY 2321 GlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsn 2340
 Db 6962 GGAACAGGCTTATGACTATAGAGCAGCGGTGAGAACTGACACCAACATCATG 7021
 QY 2341 GlyLeuGlnAlaLleLleArgLysAlaLeuMetGlyLysLysThrAspGlnTProGlu 2360
 Db 7022 GGGCTGAGGGCCATTAATTAGAAAGGCACTCATGTGTAATATACCGTGGAGAGTCC 7081
 QY 2361 ProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAla 2380
 Db 7082 CCGCCGCTCAGCCCAATGCTTTTAACTCTGAATCCAGTCCAGCTCCCGCTGCT 7141
 QY 2381 MetProLleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGly 2400
 Db 7142 ATCCCATATACCCCTGCTGACGAGCGAGTGCACACACTCATCTGCGCAGGTGGGCG 7201
 QY 2401 GlyLysAlaLysValSerGlyArgProSerSerArgLysValLysSerProAlaPro 2420
 Db 7202 GGGAGGCGCAAGGTCTGTGGAGAGCCAGAGCCGAAAGCAAGTCCCGCGCCCGG 7261

QY 2421 LeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluGlyAspCys 2440
 Db 7262 CTGGCATCTGGGAGCCGCGCACCTTGTCTCTCAATGACCTCGAGGAGACTGCAC 7321
 QY 2441 ArgArgThrProLeuThrAsnArgValTProGluAspArgProSerSerAlaGlySer 2460
 Db 7322 CGCCGAGCGCGCTCAGCAACCGGTGTGGAGGACAGGCTTGTCTCGCAGTTCAC 7381
 QY 2461 ProPheProTyrAsnProLeuLleMetArgLeuGlnAlaGlyValMetAlaSerPro 2480
 Db 7382 CCATTCCTTCAACACCTGATCATCGGCTGAGGCGGGTGTATGCTTCCCAACC 7441
 QY 2481 ProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaTPro 2500
 Db 7442 CCACCGGGCTCCCGCGGAGCGGCGCCCTGCTGCTGCCCGCCACAGCTGGGAGCAG 7501
 QY 2501 GluProLysProLeuLeuCysSerGlnTyrGluThrLeuSerAspSerGlu 2517
 Db 7502 GAGCCCAAGCCACTGCTGTCTGCACTACGAGACACTTCCAGACGAG 7552

RESULT 3

ID ADN04304 standard; cDNA; 8561 BP.

AC ADN04304;

DT 01-JUL-2004 (first entry)

DE Antipsoiatric cDNA sequence #352.

XX de; gene; antipsoiatric; gene therapy; psoriasis; diagnosis.

XX Homo sapiens.

OS W02004028479-A2.

PD 08-APR-2004.

PF 25-SEP-2003; 2003MO-US030907.

PR 25-SEP-2002; 2002US-0414006P.

XX (GENTH) GENENTECH INC.

PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WT;

DR WPI; 2004-305105/28.

XX P-PSDB; ADN04305.

PT New PRO nucleic acid or polypeptide, useful for preparing a
 PT pharmaceutical composition for diagnosing or treating psoriasis in a
 PT mammal.

XX Claim 1; SEQ ID NO 698; 3063bp; English.

CC The invention relates to novel polynucleotide and polypeptides for
 CC treating psoriasis or a sequence having at least 80% identity to the
 CC above sequences. The nucleic acid is useful for preparing a composition
 CC for diagnosing or treating psoriasis in a mammal. This sequence
 CC corresponds to one of the polynucleotides of the invention.

XX Sequence 8561 BP; 1862 A; 3033 C; 2525 G; 1141 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0 Length: 8561
 Score: 13215.00 Matches: 2517
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 12 Gaps: 0

US-09-522-753-5 (1-2517) x ADN04304 (1-8561)

QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTPaGAlaThrGluProArgTyrPro 20
DB 2 ATGTGGGCTCCACACAGCTTGGCACAGAGTGGAGGCGCACTGAGCCCGCTACCCG 61
QY 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
DB 62 CCCACACAGCTTCTTACCCAGATCGAGATCGCCGAGACACACAGCACTCGGGCTCTCG 121
QY 41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
DB 122 GAGTACACAGACACTCCCGCACTATGCTCCACCTGTGCGCGGAGCTTCATCATCCAG 181
QY 61 ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
DB 182 CCCACCGCGGAGGCGCTCCCTGCTGTGAGTTCCAGCCCGGAAATACAGCTCCAG 241
QY 81 GluLeuHisLeuArgProGlnSerHisSerTyrLeuProGluLeuGlyAsnSerGluMet 100
DB 242 GAGCTCCACCTCGGCGCAGAGTCCCATCTATCACTGCGCAGACTGGGAGATCAGAGATG 301
QY 101 GluPheIleGluSerTyrAspProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
DB 302 GAGTTCATTGAAAGCAAGCGCTCGCTGCTGAGCTGCTGACCCCTGCTGGAGCCG 361
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrTyrAspArgSer 140
DB 362 TCACCCCTGCTGGCCACCGGCGCAGCTGGGAGATCGAAGACTCACCAAGAGACGTAG 421
QY 141 LeuThrGlyLysLeuGlnProValSerProProSerProProHisThrAspProGluLeu 160
DB 422 CTGACGGGCAAGCTGGAACCGGTGTCTCCCGCAGCGCCCGCACACTGACCTTGAGCTG 481
QY 161 GluLeuValProProArgLeuSerTyrSerGluLeuIleGlnAsnMetAspArgValAsp 180
DB 482 GAGCTGTGTGCGCCACAGGCTGTCCAAAGAGAGCTGATCCAGAACTGGACCGCGTGCAC 541
QY 181 ArgGluIleThrMetValGlnGlnGlnIleSerTyrSerLysLeuLysLysGlnGlnLeu 200
DB 542 CGAGAGATCAACATGATGAGCAGCAGATCTTAAGCTGAAGAGAGAGCAGCAGCAGCTG 601
QY 201 GluGlnGluAlaAlaLysProProGluProGluLysProValSerProProIleGlu 220
DB 602 GAGGAGAGAGCTGCCAAGCGCCCGAGCTCGAAGACCCGTGTCAACCGCCGCATCGAG 661
QY 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgTyrValGlnAla 240
DB 662 TCGAAGCACCGAGCTGTGTGAGATCATCTACGACGAGAACCGGAGAGAGGCTGAGCT 721
QY 241 AlaHisArgIleLeuGlnGlyLeuGlyProGlnValAlaGluProLeuTyrAsnGlnPro 260
DB 722 GCACTCGGATTCGGAAGGCTGGGCGCCCAAGTGGAGCTGCGCTGTACAAACAGCCG 781
QY 261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
DB 782 TCCGACACACCGGACAGATCATGAGAACATCAAAATTAACACAGCGATGCGGAAAGCTA 841
QY 281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnThrLysGlnLysPheCysGln 300
DB 842 ATCTGTACTTCAAGGAGGAGATCACGCTCGAACAATGAGAGCAAGATTTGCGCAG 901
QY 301 ArgTyrAspGlnLeuMetGluAlaLeuGlnLysLysValAlaGlyIleGluAsnAsnPro 320
DB 902 CCTATGACACAGCTCATGAGGCTTGGAAAAAAGTGGAGGCGATCGAAAAACACCGG 961
QY 321 ArgArgArgAlaLysGlnSerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340
DB 962 CCGCGGCGGCGCAAGAGAGCAAGGTGCGGAGTACTTACGAAAAAGCAGTTCCCTGAGATC 1021
QY 341 ArgLysGlnArgGluLeuGlnGlnLysArgMetGlnSerArgValGlyLysArgGlySerGly 360
DB 1022 CCGAAGCAGCGGAGCTGAGAGCGCATGACAGCAGGAGGTGGCCAGCGGCGGCACTGGG 1081

QY 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
DB 1082 CTGTCCATGTGGCGCGCCGACGACGACAGAGGTGTGAGAGATCATGATGCTCTTCA 1141
QY 381 GluGlnGlnAsnLeuGlnLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400
DB 1142 GAGCAGAGAACCTTGAGAAAGAGATGCGCAGCTGGCGGTGATCCCGCCATGCTGTAC 1201
QY 401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420
DB 1202 GACGCTGACACAGACGCGATCAAGTTTCAACATGAAGAGGCGCTTATGCGCAGCCCATG 1261
QY 421 LysValTyrLysAspArgGlnValMetAsnMetTPSerGluGlnGluLysGluThrPhe 440
DB 1262 AAGGTGTACAAAGACCGCAGGTCAATGATGTGAGTGTGACAGAGAGAGAGACTTTC 1321
QY 441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
DB 1322 CCGAGAAAGTTCAATCAGATCCCAAGAACTTTGAGCTGATCGCATCTTCTGGAGAGG 1381
QY 461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480
DB 1382 AAGACAGTGGCTGAGTGGCTCTCTATTACTACCTGACTGAAGAAATGAGAACTATAAG 1441
QY 481 SerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGln 500
DB 1442 AGCTGTGTGAGACGAGAGCTATCGCGCGCGCAAGACCGACAGCAACACAGCAGCAG 1501
QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluLys 520
DB 1502 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1561
QY 521 AspGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLys 540
DB 1562 GATGAG 1621
QY 541 AspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAsnAsnProGlu 560
DB 1622 GACAAAGAGAACTCTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1681
QY 561 LysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGly 580
DB 1682 AAGAGAGCTGTGCTCCAAAGCGCCGCAAACTGCCAAAGCCAGAGAGAGAGAGAGAG 1741
QY 581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600
DB 1742 CGCATCACCCGCTCAATGCTTAATGAGGCCAAACAGAGAGGCCATACCCCCAGCAG 1801
QY 601 SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTyrThrGluGluGluMet 620
DB 1802 AGCGCCGAGCTGGCTCCATGAGCTGAATGAGATTTCTCGCTGACAGAGAGAGAGAG 1861
QY 621 GluThrAlaLysLysGlyLeuLeuGlnLysIleGlyAsnAsnTPSerAlaIleAlaProMet 640
DB 1862 GAAACAGCCAAAGAAAGGTCTCTGAAACAGCGCCCAACTGCTGCGCAATCCCCGAGATG 1921
QY 641 ValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGln 660
DB 1922 GTGGGCTCCAGACTGTGTGAGGTGAAGAACTTCTCACTCACTCAAGAAAGAGCAG 1981
QY 661 AsnLeuAspGluIleLeuGlnGlnHisLysLysLeuLysMetGluLysGluArgAsnAlaArg 680
DB 1982 AACCTCGATGAGATCTTGACAGCAGCAAGCTGGAAGATGAGAGAGAGAGAGAGAGCG 2041
QY 681 ArgLysLysLysLysAlaProAlaAlaAsnSerGluGluAlaAlaPheProProValVal 700
DB 2042 AGAAGAGAGAGAGAGAGAGCGCGCGCGCGCAGCAGAGAGAGCTGCAATCCCGCCGTGTG 2101
QY 701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGlu 720
DB 2102 GAGGATGAGAGATGAGAGCGCTCGGCGCTGAGCCGAATGAGAGAGAGATGTGTGAGAG 2161
QY 721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAla 740

Db 2162 GGTGAAGCCCTTACATGCTCTGGGAATGAGTGGCCAGAGGGGATACAGTGGCCAGCC 2221
Qy ThrValAsnAsnSerSerAspThrGluSerIleProSerProHistiThrgluAlaIleValys 760
Db 2222 ACTGTCAACAACAGCTCAGACACCGAGAGCATCCCTCTCTCTCACTGAGGGCCGCAAG 2281
Qy AspThrGlyGlnAsnGlyProLysProProAlaThrLeuGluValAspGlyProProPro 780
Db 2282 GACACAGGGCAGATGGGCGCCCAAGCCCACTCTGGGGCGCGAGGGGCCACCCCA 2341
Qy 781 GlyProProThrProProAlaGlyThrSerArgAlaProIleGluProThrProAlaSer 800
Db 2342 GGCCCAACCCACCCACACGAGAGACATCCCGGGCCCCCATTTGAGCCACCCGGCCTCT 2401
Qy GluAlaThrGlyValaProThrProProProAlaProProSerProSerAlaProProPro 820
Db 2402 GAAGCCACCGGAGCCCTTACGCCCCCACAGCACCCCATGCTGCTGCACTCTCTCT 2461
Qy 821 ValValProLysGluGluGlyGluGluThrAlaAlaAlaProProValGluGluGly 840
Db 2462 GTGTGCCCCAAGAGAGAGAGAGAGAGAGACCGGACAGCCGCCAGTGGAGAGGGG 2521
Qy 841 GluGluGluGluLysProProAlaAlaGluGluLeuAlaIleThrGlyValaGluGlu 860
Db 2522 GAGGACAG 2581
Qy 861 ProValLysSerGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 880
Db 2582 CCGGTCAAG 2641
Qy 881 GluAlaAlaGluAlaThrAlaGluGluGluAlaLeuLysLeuGluGluGluGluGlySer 900
Db 2642 GAGGCGCGTGAAGGCGAGGCGGCGGCGCTCAAGCGAGAGAGAGAGAGAGAGAGAG 2701
Qy 901 GlyArgAlaThrThrAlaLysSerSerGlyAlaProGluAspSerAspSerSerAlaThr 920
Db 2702 GGCAGGCGCACCACTGCAAGAGCTGGGGGCCCCCAGAGAGAGAGAGAGAGAGAG 2761
Qy 921 CysSerAlaAspGluValaIleAspGluAlaGluGluGluGluGluGluGluGluGlu 940
Db 2762 TGCAGTCAAGACAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2821
Qy 941 ArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGluLysPro 960
Db 2822 AGGCCCAAGCTCTCACTCCGAGCTGGGAGACCCCGGGCCATCTCACTCCAGAGAGCA 2881
Qy 961 LeuAspLeuLysGluLeuLysGluAlaAlaIleProProIleGluValaThrLys 980
Db 2882 CTGGACCTGAAGAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2941
Qy 981 ValHisGluProProArgGluAspAlaAlaProThrLysProAlaProAlaProPro 1000
Db 2942 GTCCATGAGCCCGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3001
Qy 1001 ProProGluAsnLeuGluProGluSerAspAlaProGluGluProGluSerProArg 1020
Db 3002 CCACCGCAAAACCTGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3061
Qy 1021 GlyLysSerArgSerProAlaProProAlaAspLysGluAlaAlaGluAlaGlu 1040
Db 3062 GGCAG 3121
Qy 1041 LysLeuProGluLysAspProProCysTyrThrSerGlyLeuProThrProAlaProPro 1060
Db 3122 AGCTGCTGCTGGAG 3181
Qy 1061 GluValIleLysAlaSerProHisAlaProAspProSerAlaAspSerTyrAlaProPro 1080
Db 3182 GAGGTATATCAAGGCTCCCGGAGATGGCCCGAGAGAGAGAGAGAGAGAGAGAGAGAG 3241
Qy 1081 GlyHisProLeuProLeuGluGluLysLeuHisAspThrAlaArgProValLeuProArgProPro 1100

Db 3242 GGTCAACCACTGCGCTGGGCTCCATGACATGCCCGGCGGCTCTCTCCGCGCCCAACC 3301
Qy ThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeuGluAlaArg 1120
Db 3302 ACCATCTCCAAACCGCTCCCTCTCATCTCTGTGCAAGAGAGAGAGAGAGAGAGAGAG 3361
Qy 1121 GlnIleGlyAlaIleSerGluGlyMetSerValGlnLeuHisValProTyrSerGluHis 1140
Db 3362 CAATATGATCCATCTCCCAAGAGATGTGGTCCAGTCCAGTCCCGGTACTCAGAGCAT 3421
Qy 1141 AlaLysAlaProValGlyProValThrMetGlyLeuProLeuProPoweAspProLys 1160
Db 3422 GCCAAGGCGCGGTGGGCTGTACCATGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3481
Qy 1161 LeuAlaProThrSerGlyValLysGluGluGluGluGluGluGluGluGluGluGlu 1180
Db 3482 CTGGACCTTTCAGCGAG 3541
Qy 1181 ProGluSerLeuGlyValProThrAlaGluGluAlaSerValLeuArgGlyThrAlaLeu 1200
Db 3542 CCGGAGAGCTGGGGGGTGGCCACAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3601
Qy 1201 GlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAsp 1220
Db 3602 GGCTCAGTTCCGGGCGGAGAGATCACCAAGGATTCACAGAGAGAGAGAGAGAGAGAG 3661
Qy 1221 SerAlaIleThrThrArgGlySerIleThrHisGlyThrProAlaProValLeuTyrLys 1240
Db 3662 AGCGCATCAATACCGGCTCATACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3721
Qy 1241 GlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgIleArgGlu 1260
Db 3722 GGCACATACCAAGATCATCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3781
Qy 1261 AspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyr 1280
Db 3782 GACAGCTTCCCAAGGGCGACATCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3841
Qy 1281 GluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyValArgSerSerGlyPro 1300
Db 3842 GAGGAGGAGAGTGTCTGTACCCAGTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3901
Qy 1301 ProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArg 1320
Db 3902 CCCCATGAG 3961
Qy 1321 AlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHis 1340
Db 3962 GCCATCTCTCAGCAGATCGAAGGCTCATGGGCGGTGCCATCCCGCGGAGAGAGAG 4021
Qy 1341 SerProHisIleLeuLysGluGluGluGluGluGluGluGluGluGluGluGluGlu 1360
Db 4022 AGCCCCACACCTCAAAAG 4081
Qy 1361 ArgSerTyrValGluAlaGluGluAspTyrLeuArgArgGluAlaLysLeuLeuLysArg 1380
Db 4082 CGGTCTTATGAG 4141
Qy 1381 GluGlyThrProProProProProProProSerArgAspLeuThrGluAlaTyrLysThrGln 1400
Db 4142 GAGGAG 4201
Qy 1401 AlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValaIleThrValaLysGlu 1420
Db 4202 GCGTGGAG 4261
Qy 1421 AlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuPro 1440
Db 4262 GCGGGGCGGTCCATCATGATATCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4321
Qy 1441 LeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyrAsp 1460
Db 4322 CTGGCCCGCGGCGCTCAAG 4381

QY 1461 ThrGlyAlaSerThrThrglySerlyshisAspValArgSerLeuileGlySerPro 1480
DB 4382 ACCGGCGGCTCCACCACTGGCTCAAAAAGCAGAGTAGCTCCCTCATCGCGAGCCCC 4441
QY 1481 GlysArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGlu 1500
DB 4442 GGCCTGACCTTCCACCCCTGACCCCGTAGATGTATGCGGACCGCCGGGACACTGGAA 4501
QY 1501 ArgAlaCyTYrGlyGluSerLeuysSerArgProGlyThrAlaSerSerSerGlyGly 1520
DB 4502 CCGTGGCTGTAAGAGAGAGAGCTGAAGAGCCGGCCAGGAGCCGCCAGAGCTCGGGGGCC 4561
QY 1521 SerileAlaArgGlyAlaProValileValProGluLeuGlyLysProArgLysSerPro 1540
DB 4562 TCATATTGGCGCGGGCGCCCGGTGATTGTGCTGAGCTGGGTAAAGCCGGCGAGAGCCCC 4621
QY 1541 LeuThrTYrGlyLysAspHisGlyAlaProPheAlaGlyHisLeuProArgLysSerProVal 1560
DB 4622 CTGACCTATGAGGAGACACCGGGGACCTTTGGCGGACCTCCACGAGAGTTGCGCCGTG 4681
QY 1561 ThrMetArgGluProThrProArgLeuGluGlySerLeuSerSerSerlysaSer 1580
DB 4682 ACCATCGGGGAGCCCAAGCCGGGCTGCGAGAGGAGCCTTTCTGTCAGCAAGGATCC 4741
QY 1581 GluAspArgLysLeuThrSerThrProArgGluileAlaLysSerProHisSerThrVal 1600
DB 4742 CAGGACCGGAAGCTGACGTGAGCGCTCGTAGATCGCAAGTCCCGCAGCAGCACCGTG 4801
QY 1601 ProGluHisHisProHisProIleSerProTYrGlyHisLeuLeuArgLysValSerGly 1620
DB 4802 CCGGAGCACACCCACACCCCATCTGCGCTATGACACCTGCTGGGGCGGTGATGTGCG 4861
QY 1621 ValAspLeuTYrArgSerHisileProLeuAlaPheAspProThrSerileProArgGly 1640
DB 4862 GTGGACCTTATCGCGACCATCCCCCTGGCTTGACCCCACTCCATACCCCGGGC 4921
QY 1641 IleProLeuAspAlaAlaAlaAlaTYrTYrLeuProArgHisLeuAlaProAsnProThr 1660
DB 4922 ATCCCTCTGAGCGAGCCGCTGCTACTGACTGACCTGCGCAGACCTGGCCCCCAACCCACCC 4981
QY 1661 TYrProHisLeuTYrProProTYrLeuileArgGlyTYrProAspThrAlaAlaLeuGlu 1680
DB 4982 TACCCGCACTGTACCCACCTACCTACCTCGCGGTACCCCAACCGCGGCGGTGAG 5041
QY 1681 AsnArgGluThrileIleAsnAspTYrileThrSerGluGlnMetHisHisAsnThrAla 1700
DB 5042 AACCGGACAGCATTCATGACTGACTGACTGCTCGCAGCAGATGACCAACAGAGGCC 5101
QY 1701 ThrAlaMetAlaGlnArgAlaAspMetLeuArgLysLeuSerProArgLysSerSerLeu 1720
DB 5102 ACCGCGCATGGCCAGGAGCTGATGTCGAGGGGCGCTCGGCCCGCAGAGTCTCGCTG 5161
QY 1721 AlaLeuAsnTYrAlaAlaGlyProArgGlyileIleAspLeuSerGlnValProHisLeu 1740
DB 5162 GAACCTCACTAGCGTCGGGTCCCGAGGACATGATGACCTGTCCCAAGTGCACACCTG 5221
QY 1741 ProValLeuValProProThrProGlyTYrProAlaThrAlaMetAspArgLeuAlaTYr 1760
DB 5222 CCGTGTCTGTGCCCCCGGACACGAGCAACCCAGCCAGCCCATGAGACCGCTTGCTGCTAC 5281
QY 1761 LeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGly 1780
DB 5282 CTCGCCACCGCGCCCGAGCCCTTCAGACGCGCCAGCAGAGCTCCCACTCTCCAGGA 5341
QY 1781 GlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGlyLysAspArg 1800
DB 5342 GGCTCAACACACTTGCAAAACCAACACACCACTCTCGTCCGAGGGGAGCAGACCGG 5401
QY 1801 AspArgGluArgAspArgAspArgLysGlyLysSerileLeuThrSerThrThrThr 1820
DB 5402 GATCGAGAGCGGAGCCGGGATCGGAGCGGGAAGTTCATCTTCACTGTCACCAAGACG 5461

QY 1821 ValGluHisAlaProIleThrArgProGlyLysThrgluGlnSerSerGlySerSerGlySer 1840
DB 5462 GTGGAGCAGCACCCATCTGGAGACTGTATCAGAGAGAGAGCGGCGAGCGGAGCC 5521
QY 1841 SerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHis 1860
DB 5522 ACCGGCGGGGTGGGGGACGACGAGCCGCCCGCTCCCACTCCCATGCCACCAACGAC 5581
QY 1861 SerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsn 1880
DB 5582 TCGCCCATCTCCCTCGGAGCCAGAGATGCCCTCCAGAGAGACCAAGTGTGCTTCAAC 5641
QY 1881 ThrGlyMetLysGlyileIleThrAlaValGluPheProSerLysProThrAlaLeuArgSer 1900
DB 5642 ACAGGATGAAAGGATATCATCGCTGTGAGGCCAGAACCCCAAGGCTCTGAGGTCC 5701
QY 1901 ThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCyPro 1920
DB 5702 ACCTCCACTCTCTCAACCGCTTGCCAGCTGCCACATTCGCCACCTGCCACCGCCCA 5761
QY 1921 LeuGlyGlyThrLeuAspGlyValTYrTYrProThrLeuMetGluProValLeuLeuProLys 1940
DB 5762 CTGGCGGGACCCCTGATGGGGTTACCTTACCTCATGAGACCGGTCTTGCTGCGCAAG 5821
QY 1941 GluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLeu 1960
DB 5822 GAGGGCCCCGGGTGCGCCGCGCAGAGCGGCCCGCAGCAGACACCGCCATCGCTTCTTC 5881
QY 1961 AlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerlysglySerGlu 1980
DB 5882 GCCAAGCCCCCAGCCCGCTCGGGGTGAGGCCGCGCTCTCCCGCAGCAAGGGGCTCGAG 5941
QY 1981 ProArgProLeuValProProValSerGlyHisAlaThrileAlaArgThrProAlaLys 2000
DB 5942 CCGCGGCCCTGATGCGCTCTGCTCTGCGCAGCAGCATGAGCCCGCAGCCCTCGGAAG 6001
QY 2001 AsnLeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSerAsp 2020
DB 6002 AACCTCGCACTTACCAACCGCAGCCGAGCCCGCGCGCGCACCTGCTCGGCTCGGAC 6061
QY 2021 ProHisArgGluLysThrGlnSerLysProPheSerileGlnGluLeuGluLeuArgSer 2040
DB 6062 CCGCAGCGGAAAGACTCAAGTAAACCTTTTCATCATCAGGAAGTGAACCTCGCTTCT 6121
QY 2041 LeuGlyTYrHisGlySerSerTYrSerProGluGlyValGluProValSerProValSer 2060
DB 6122 CTGGGTTACACGCGGAGCAGTACAGCCCGCAAGGGGTGAAGCCCTGACCTGTGAGC 6181
QY 2061 SerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSer 2080
DB 6182 TCACCAAGTCTTACCCACGACAAAGGGGCTCCCAAGCACCCTGGAAAGCTCGACAAAGAGC 6241
QY 2081 HisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGlyAla 2100
DB 6242 CACCTGAGAGGGAGCTGGCGGCCAAGCAGCCAGCGCTTGAAGCTTGCGGGGAGGCC 6301
QY 2101 AlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSerProLeuLeu 2120
DB 6302 GCCCACTCCACACCTGGCGCGCTGCTGAGAGCAGCACTTCGTCCAGCCCGCTGCTC 6361
QY 2121 GlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSer 2140
DB 6362 CAGACCGCCCGAGGGGTCAAAAGTCAACGCGGGTGTACCTTGCCCAACATCAAG 6421
QY 2141 GluValileThrGlnAspTYrThrArgHisHisProGlnGlnLeuSerAlaProLeuPro 2160
DB 6422 GAGGTATATCACAGGACTACACCGGACACACCAACAGAGTACGCGCAACCTCGGCC 6481
QY 2161 AlaProLeuTYrSerPheProGlyAlaSerCyProValLeuAspLeuArgArgProPro 2180
DB 6482 GCCCCCTCTACTCTTCCCTGGGGCAGCTGCCCGCTGGAACCTCGCGCCGACCC 6541
QY 2181 SerAspLeuTYrLeuProProProAspHisGlyAlaProAlaArgLysSerProHisSer 2200

Db	6542	AGTGA	CGCTAC	CTACCTCC	CCCCCCCCGGGACCA	TGCTGCCGCGCGCTCC	CCCCCA	CACG	6601
Qy	2201	GIUGI	GYLYV	YAS	SerProGIuP	ProAnlyS	ThrSerVal	LeuGIYGIYGIuA	2220
Db	6602	GAA	GGGGGCA	AGAGCT	CTCCAGAGCA	AAAGAGTGGCTTGGTGGTGA	GAGAC		6661
Qy	2221	GIYI	LegIuP	ProValSer	ProProGIuGI	MetThrgIu	ProGIuHISer	ArgSerA	2240
Db	6662	GGA	TATTA	AACTCTGT	CCCCCA	CGGAGGACAT	AGCCAGGCACT	CCCCAGTCT	6721
Qy	2241	ValY	TrProLeu	LeuYTr	ArgA	AspGIYGIu	InThrgIu	ProSerArgMetGI	2260
Db	6722	GTG	TAC	CGCGCTGT	ATACCGGGA	TGGGAA	CAGACG	AGCCAGAGATGGCT	6781
Qy	2261	SerPro	GIuA	AntThrSer	GInPro	ProBla	A	PheB	2280
Db	6782	TCT	CAGGCA	ACACCG	CAGCGCACCT	TTTCAC	AAAGCTGA	CCGGAAGCACTCC	6841
Qy	2281	Ala	MetVal	Y	SerSerY	LYSGIn	GIuIle	AnlyS	2300
Db	6842	GCA	TATG	TCACAGT	CCCAAGAC	AGATCA	CAAGAGCT	GAACCCCA	6901
Qy	2301	GIu	ProGIu	YTrA	AsnIle	SerGIu	ProGIu	YThrgIuIle	2320
Db	6902	GAG	CTGA	TAATCA	TATACG	CAGCTTGGG	ACGGAATCTTCA	TATATCCG	6961
Qy	2321	GIY	ThrgIu	LeuMetThr	YTr	ArgSerGIu	AlaValGIu	InHISerTh	2340
Db	6962	GGA	ACAGGCT	TATAGACCT	TATAGAAC	CGAGCGGTGC	AGAACATG	CCAGCAAC	7021
Qy	2341	GIY	LeuGIu	AlaIleIle	ArgY	ValAla	LeuMetGIY	STY	2360
Db	7022	GCG	CTGAGG	CCATAT	TATGAAG	GCATCTAG	GGTAAAT	TATGAC	7081
Qy	2361	Pro	ProLeu	SerAla	AsnAla	Phe	AsnPro	LeuAsnAla	2380
Db	7082	CGG	CGCTAC	AGCGCA	AAAGCTTT	TAAACCTT	GTAAATG	CAATGTCC	7141
Qy	2381	Met	ProIle	ThrAla	Ala	AspGIY	ArgSer	AspHIS	2400
Db	7142	ATG	CCCA	TAAACG	GTGTG	ACGACG	AGATCA	CAACACTC	7201
Qy	2401	GIY	Y	ValAla	ValSerGIY	ArgPro	SerArg	ValAla	2420
Db	7202	GGA	AGAG	CCAAAG	GTCTCTG	GCAGAC	CCGACG	CCGAAAGCA	7261
Qy	2421	Leu	Ala	SerGIY	AspArgPro	ProSer	ValSerSer	ValHISerGIu	2440
Db	7262	CTG	GCAT	CTGGG	ACCGG	CCACCTCT	GTCTCTCA	GTGCACTG	7321
Qy	2441	Arg	Arg	ThrTrPro	LeuThr	AsnArg	ValITrgIu	AspArgPro	2460
Db	7322	CGC	CGGAC	CGCGCT	ACCAACCG	GTGTGG	AGGAGAC	AGGCGCTGT	7381
Qy	2461	Pro	Pro	Pro	YTrA	AsnPro	LeuIle	MetArg	2480
Db	7382	CCA	TTTCCC	CTTACAC	CCCCCTG	ATATATG	GGGTG	AGGAGTCA	7441
Qy	2481	Pro	Pro	GIu	YLeu	ProAla	GIYSerGIY	ProLeu	2500
Db	7442	CCA	CGGG	CGCTCC	CGGGG	ACGCGG	CGCTG	CTGGCCCC	7501
Qy	2501	GIu	Pro	Y	AsnPro	LeuLeu	CY	AsnSerGIu	2517
Db	7502	GAG	CCCA	AGCCAC	ATGCTCTG	CTGCAG	TAACAG	ACATCTTCC	7552

08-FEB-2001 (first entry)

Human ORFX ORF338 polynucleotide sequence SEQ ID NO:675.

Human; open reading frame; ORFX; detection; cytosolic; hepatotropic; viremia; antiproliferative; antiparkinsonian; neurotropic; neuroprotective; antiviral; osteopathic; antiarthritic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antilinfamatory; antiviral; antibacterial; antitungal; antirheumatic; antihypertoid; antianemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive; ss.

Homo sapiens.

WO200058473-A2.

05-OCT-2000.

31-MAR-2000; 2000MO-US008621.

31-MAR-1999; 99US-0127607P.
02-APR-1999; 99US-0127636P.
05-APR-1999; 99US-0127728P.
30-MAR-2000; 2000US-00540763.

(CUBA-) CUBAGEN CORP.

Shimkets RA, Leach M,
WPI; 2000-602362/57.
P-PSDB; AAB40574.

Novel nucleic acids and peptides derived from open reading frame X, useful for treating e.g. cancers, proliferative disorders, neurodegenerative disorders and cardiovascular disease.

Claim 5; Page 779-784; 5507pp; English.

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43387, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytostatic; hepatotropic; viremia; antiproliferative; antiparkinsonian; neurotropic; neuroprotective; osteopathic; anticonvulsant; antiarthritic; immunosuppressant; immunostimulant; cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antibacterial; antiviral; antitungal; antirheumatic; antihypertoid; and antianemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX-associated disorder. The nucleic acids can be used to express ORFX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive

Sequence 8564 BP; 1859 A; 3034 C; 2532 G; 1139 T; 0 U; 0 Other;

RESULT 4	
AAC74783	
ID	AAC74783 standard; cDNA; 8564 BP.
XX	
AC	AAC74783;
XX	

Alignment Scores:	
Pred. No.:	0
Score:	13178.50
Percent Similarity:	99.80%
Best Local Similarity:	99.76%
Length:	8564
Matches:	2512
Conservative:	1
Mismatches:	4

Query Match: 99.72% Indel:s 1
DB: 3 Gaps: 1
US-09-522-753-5 (1-2517) x AAC74783 (1-8564)

QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTTPArgAlaThrGluProArgTyrPro 20
DB 1 ATGTCCGGGGCTCCACACAGCTTGTGGCAGACCTGGAGGGCCACTAGAGCCCCCTCCACCG 60
QY 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
DB 61 CCCACAGCCCTTTCTTACCCAGTGCAGATCGCCGAGCGACAGAGCGTCCGGCTCTCG 120
QY 41 GlnTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
DB 121 GAGTACACAGACACACTCCCGGACGTATGCTCCACCTGTCCGGGGCTTCATCATCCAG 180
QY 61 ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyValenGluArgSerGln 80
DB 181 CCCACAGCGCGAGGGCCCTCCCTGCTGTCTGAGTTCAGCCCGGGAATGAACGGTCCAG 240
QY 81 GlnLeuHisLeuAspArgProGlnSerHisSerTyrLeuProGlnLeuGlyLysSerGlnMet 100
DB 241 GAGCTCACCTGGGCGACAGATCCCACTATACCTGCGAGCTGGGGAAGTCAGAGATG 300
QY 101 GlnPheIleGlnSerLysArgProArgLeuGlnLeuLeuProAspProLeuLeuArgPro 120
DB 301 GAGTTCATTGAAAGCAAGAGCGCCCTCGGCTAGAGCTGCTGCTGACCCCTGCTGCGAGCG 360
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
DB 361 TCACCCCTCTGCGCCACGGGCGCACCTGGCGGATCTGAAGACCTTCACAAAGACCTTACG 420
QY 141 LeuThrGlyLysLeuGlnProValSerProProSerProProHisThrAspProGlnLeu 160
DB 421 CTGACGGGAGAGCTGGAACCGGTGTCTCCCCAGGCCCGCCGACACTGACCTGAGCTG 480
QY 161 GlnLeuValProProArgLeuSerLysGlnGlnLeuIleGlnAsnMetAspArgValAsp 180
DB 481 GAGCTGTGTCCCGCACGGCTGTCCAAAGAGAGCTATCCAGAACATTCAGACCGCGTGGAG 540
QY 181 ArgGlnIleThrMetValGlnGlnGlnIleSerLysLeuLysValLysGlnGlnLeu 200
DB 541 CCAGAGATCATCCATGCTAGAGCAGCAGATCTTAACTTAAAGAAAGCAGCAACGACTG 600
QY 201 GlnGlnGlnAlaAlaLysProArgProGlnProGlnLysProValSerProProIleGln 220
DB 601 GAGGAGAGAGCTGCCAAGCCGCCGAGCCTGAGAAAGCCCGTGTACCCGCCCATCGAG 660
QY 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGlnAsnArgLysLysAlaGlnAla 240
DB 661 TCGAAGCACCGCAGCCTGTGTGCAGATCATCTACGAGAGAACCGGAAGAGCTGAAGCT 720
QY 241 AlaHisArgIleLeuGlnGlnLysGlnProGlnValGlnLeuProLeuTyrArgAsnGlnPro 260
DB 721 GCACATCGGATTTCTGAAAGCCTCGGGGCCCAAGTGAAGTCCGCTGTATCAACACGACC 780
QY 261 SerAspThrArgGlnTyrHisSerLysAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
DB 781 TCCGACACCCCGGAGATCATGAGAACATCAAAATAAACAGCGATGCGGAAGAGCTA 840
QY 281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGln 300
DB 841 ATCTTGTACTTCAAGAGAGAGATCAAGCTCGAAGAACATGGAGAGCAAGATTCTGCCAG 900
QY 301 ArgTyrAspGlnLeuMetGlnAlaLeuGlnLysLysValGlnArgIleGlnAsnAsnPro 320
DB 901 CGCTATGACACAGCTCATGAGGCTCGGGAAGAAAAAGTGAAGCGCATTCGAGAACACCCC 960
QY 321 ArgArgArgAlaLysGlnSerLysValArgGlnTyrTyrGlnLysGlnPheProGlnIle 340
DB 961 CGCGCGCGGGCCCAAGAGAGCAAGAGTGCAGAGTACTTACGAAAAAGAGATTCCCTGAGATC 1020

QY 341 ArgLysGlnArgGlnLeuGlnGlnLysMetGlnSerArgValGlyGlnArgGlySerGly 360
DB 1021 CCCAAGCAGCGGAGCTGCAGAGAGCCATAGCAGGCGGCTGGCCAGCGGGGCAAGTGGG 1080
QY 361 LeuSerMetSerAlaAlaArgSerGlnHisGlnValSerGlnIleIleAspGlyLeuSer 380
DB 1081 CTGTCCATGTGGAGCCCGCCGACGAGCAGAGAGTGTCAAGATTCATCATGAGCTCTCA 1140
QY 381 GlnGlnGlnLysLeuGlnLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400
DB 1141 GAGCAGAGAACCTGAGAGAGAGAGATGCCGACGTGGCCGTATCCCGCCATGCTGTAC 1200
QY 401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420
DB 1201 GAGCTGACACACAGCGCATCAAGTTCATCAATGAACGGGCTTATGGCCGACCCCAAG 1260
QY 421 LysValTyrLysAspArgGlnValMetAsnMetTyrSerGlnGlnLysGlnTyrPhe 440
DB 1261 AAGGTGTACAAAGACCGCAGGTCATGAACATGTGAGTGAAGAGAGAGAGACCTTC 1320
QY 441 ArgGlnLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGlnArg 460
DB 1321 CGGAGAGATTCATGAGCATCCCAAGAACTTGGCTGATGGCATCTTCCTGGAGAGG 1380
QY 461 LysThrValAlaGlnCysValLeuTyrTyrTyrLeuThrLysLysAsnGlnLysTyrLys 480
DB 1381 AAGACAGTGGCTGAGTGCCTCTTACTTACTGACTGACTGAAGAAATGAGAACTTAAG 1440
QY 481 SerLeuValArgArgSerTyrArgArgArgGlyLysSer---GlnGlnGlnGlnGln 499
DB 1441 AGCTGTGTAGAGAGAGCTATCGCGCCGCGAAGAGCGACAGCAGCAACACGACG 1500
QY 500 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 519
DB 1501 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1560
QY 520 LysAspGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGln 539
DB 1561 AAAGATGAGAAAGAGAAAGAAAGAGCGGAGGAGAAAGAGAGAGAGAGAGCGGAGTGGAG 1620
QY 540 AsnAspLysGlnAspLeuLeuLysGlnLysThrAspAspThrSerGlyLysAspAsnAsp 559
DB 1621 AACGACAAAGAAAGACTCTCTCAAGAGAGAGAGAGAGACACTTCAGGGAGAGCAACGAG 1680
QY 560 GlnLysGlnAlaValAlaSerLysGlnArgLysThrAlaAsnSerGlnLysArgArgLys 579
DB 1681 GAGAGAGAGGCTGTGCTCTCCAAAGCCCGCAAACTGCCAAACAGCAGGAGAGACCCAA 1740
QY 580 GlyArgIleThrArgSerMetAlaAsnGlnAlaAsnSerGlnGlnAlaIleThrProGln 599
DB 1741 GGCCTGCATCCCGCTCAATGGCTATAGAGGCCAACACAGGAGAGGCAATCACCCCCAG 1800
QY 600 GlnSerAlaGlnLeuAlaSerMetGlnLeuAsnGlnLysSerArgTyrThrGlnGlnGln 619
DB 1801 CAGAGCGCCGAGCTGCTCTCATGAGCTGAATGAGATTCCTGCTGACAGAAAGAGAA 1860
QY 620 MetGlnThrAlaLysLysGlyLeuLeuGlnHisGlyArgAsnTyrPheAsnTyrLysLysArg 639
DB 1861 ATGGAACACGCCAAGAAAGGTCTCTTGAAACACGGCCGCAATGTGTGGCCCATCGCCCG 1920
QY 640 MetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArg 659
DB 1921 ATGGTGGCTCCAAAGCTGTGTGCGAGTGAAGAACTTCACTTCAATCAAGAAAGAGG 1980
QY 660 GlnAsnLeuAspGlnIleLeuGlnGlnHisLysLysLeuLysMetGlnLysGlnLysArgAsnAla 679
DB 1981 CAGAACTTCGATGAGATCTTGTGACAGCAGCAAGCTGAAATGAGAGAGAGAGAGAGCGG 2040
QY 680 ArgArgLysLysLysLysLysAlaProAlaAlaAlaSerGlnGlnAlaAlaPheProProVal 699
DB 2041 CGAGGAGAAAGAAAGAAAGCGCGGGGCGCGCAGGAGAGAGGCTGCAATTCGCCGCCGTG 2100
QY 700 ValGlnAspGlnGlnMetGlnAlaSerGlyValSerGlyAsnGlnGlnGlnMetValGln 719

[illegible]

Db	3181	CGTAGAGTGATCAAGAGCTCCCCGGCATGCCCCGGAGACCCCTTACGGCTTCTCTACGCTTCCA	3240
QY	1080	ProGlyHisProLeuProLeuGlyYLeuHisapThraIaaArgProValLeuProArgPro	1099
Db	3241	CGTGATCAACCACTGAGCCCTGGAGCTTCATGACACTGGCCGGCCGATCTCGCGAGCCA	3300
QY	1100	ProThrIleSerAsnProProProLeuIleSerSerAlaYHisProSerValLeuGlu	1119
Db	3301	CCCAACCACTCCAACCCCGCCCTCCATCTCTCTGCAAGACACCCCAAGGCTCTCAAG	3360
QY	1120	ArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGlu	1139
Db	3361	AGCGAAATAGGTGCCATTCCTCCAAAGAAATGTCGGTCCAGCTTCCAGTCCGTACTCAGAG	3420
QY	1140	HisAlaYsaIaProValGlyProValThrMetGlyLeuProLeuProMetAspProlys	1159
Db	3421	CATGGCAAGGCCCGGGTGGGCTCGTGCATCATGGGGGTGGCCCTGCCATGACCCCAA	3480
QY	1160	LysIleuAlaProPheserGlyValIlysgGlnGlnLeuSerProArgGlyGlnIlaGly	1179
Db	3481	AAGCTGGACCCCTTCAGCGGAGTAAAGACAGACGAGCTGTCCCACGGGGCGACGGCTGGG	3540
QY	1180	ProProGlnIleSerLeuGlyValProThrAlaGlnGlnIlaSerValLeuArgGlyThrAla	1199
Db	3541	CCACCGGAGACCTCGGGGGGTGCCCCACAGCCAGAGGGGTGCGTGTAGAGGGACAGCT	3600
QY	1200	LeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValProSer	1219
Db	3601	CTGGGCTCAGTCCGGGCGGAAGCATCCAAAGCATTCACACACACGGGTCCCTCG	3660
QY	1220	AspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProIlaAspValLeuTyr	1239
Db	3661	GACAGCGGCATCATATACCGGGGCTCCATCCACACGGCACGCAAGCTGACGTCTTGAC	3720
QY	1240	LysGlyThrIleThrArgIleIleIleGlyGlnAspSerProSerArgLeuAspArgGlyArg	1259
Db	3721	AAGGCACTCATACAGAGATCATGCGGAGGACAGCCGAGTCCCTTGAGACGGGGCGGG	3780
QY	1260	GluAspSerLeuProLysGlyHisValIleTyrGlnGlyLysLysGlyHisValLeuSer	1279
Db	3781	GAGGACAGCCTCGCCCAAGGGCAGCTCATCTACGAAGGCAAGAGGGCCAGCTTGCTCC	3840
QY	1280	TyrGlnGlyGlyMetSerValThrGlnCysSerLysGlnAspGlyArgSerSerSerGly	1299
Db	3841	TATAGAGGTGCATGTCTGTGACCAAGGCTCCAAAGAGGACGGCAAGACAGCTTCAGGA	3900
QY	1300	ProProHisGlyUthrIlaIlaIlaProLysArgThrTyrThrMetMetGlnArgValGly	1319
Db	3901	CCCCCCCATGAGAGGGCGCCGCCCAAGCGCACTATGACATGATGAGGGCGCGGTGGCG	3960
QY	1320	ArgAlaIleSerSerAlaSerIleGlnGlyLeuMetGlyArgAlaIleProProGluArg	1339
Db	3961	AGAGCATCTCTCAAGCAGCATGAAGAGTCTCAATGGCCGTGCCATTCGCCGAGAGCA	4020
QY	1340	HisSerProHisIleLeuLysGlnGlnHisIleHisIleArgGlySerIleThrGlnGlyIle	1359
Db	4021	CACAGCCCCCAACCACTCAAGACAGACCAACATCCGGGGTGCATACACAAAGGATC	4080
QY	1360	ProArgSerTyrValGlnIlaGlnGlnAspTyrLeuArgArgGlnAlaYleuLeuLys	1379
Db	4081	CCTGGGTCTTACGTGAGAGGCAACAGAGAGATCACTGGGTGGGGGCGCAACTCTTAAG	4140
QY	1380	ArgGlnGlyThrProProProProProProSerArgAspLeuThrGlnAlaTyrLysThr	1399
Db	4141	CGGAGAGGACAGCTTCGCGCCCCACCGGCTCAGCGGAGCTGACCGAGGCTTACAAAGC	4200
QY	1400	GlnAlaLeuGlyProLeuLysLeuLysProAlaHisGlnGlyLeuValAlaThrValLys	1419
Db	4201	CAGGCCCCGTGGCCCCCTGAAGCTGAAGCCGGGCCCATGAGGGGCTGTGTGGCCACGCTGAAG	4260
QY	1420	GlnAlaGlyArgSerIleHisGlnIleProArgGlnGlnLeuArgHisThrProGlnLeu	1439
Db	4261	GAGCGGGGCGCTTCATCATGAGATCCCGGCGAGAGGCTGCGGCAACGCCCGGAAGCTG	4320

QY 1440 ProLeuAlaProArgProLeuLysGlySerIleThrGlnGlyThrProLeuLysTYR 1459
DB CCCCTGGCCCGCGCGCGCTCAAGAGAGGCTCATCAGGAGCAACCCGCTCAAGTAC 4380
QY 1460 AspThrGlyAlaSerThrThrGlySerIleLysValAlaArgSerIleuIleGlySer 1479
DB GAAACCGGCGCGCTCAACCTGGCTCCAAAGACAGACGTAAGCTCCCTCATCGGACAG 4440
QY 1480 ProGlyArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeu 1499
DB CCGGCGCGAGAGCTTCCACCCCGTGCACCCGCTGATGTATGACCGACGCGCGGACCTG 4500
QY 1500 GluArgAlaCysTyrGlnGlySerLeuLysSerArgProGlyThrAlaSerSerSerGly 1519
DB GAAACGCTCGCTACAGAGAGAGCTGAAGAGCCGCGCAGAGGACCGCAGACACTCGGG 4560
QY 1520 GlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysProAlaGlnSer 1539
DB GGCTCATTTGGCGCGCGCGCGCTGATTTGTGCTGAAGTGGGTAAAGCCGCGAGAGC 4620
QY 1540 ProLeuThrThyGlyLysAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerPro 1559
DB CCCCTAACTATAGAGACCAAGGAGCACTTTGCGGCGCACTCCACGAGGTTGCCCC 4680
QY 1560 ValThrMetArgGluProThrProArgLeuGlnGlySerLeuSerSerSerIleVal 1579
DB GTGACCAACCGGAGCGCCAGCGCGCTGCAAGAGGAGGAGCCTTTCGTCACAGAGCA 4740
QY 1580 SerGlnAspArgLysLeuThrSerThrProArgGlnIleAlaLysSerProHisSerThr 1599
DB TCCCAAGACCGAAAGGTGAGTGAAGCTGTGAAGATGCAAGTCCCGCACACACACC 4800
QY 1600 ValProGlnHisIleHisProHisProIleSerProTyrGlnHisLeuLeuArgGlyValSer 1619
DB GTGCCCGAGAGCAACCAACCCCATCTGCGCTTAAGAGCACTTGCTGGGCGGTGAGT 4860
QY 1620 GlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArg 1639
DB GGGGTGAGCTGTATCGAGAGCAATCCCGCTGCTGCAACCCCATCATACACCCG 4920
QY 1640 GlyIleProLeuAspAlaAlaAlaAlaIleTyrTyrLeuProArgHisIleLeuAlaProAlaPro 1659
DB GGCATCTCTGAGACCAAGCGCGCTGCTATCTACCTGCCGACCACTGGCGCCCAACCC 4980
QY 1660 ThrTyrProHisLeuTyrProTyrLeuIleArgGlyTyrArgProAspThrAlaAlaLeu 1679
DB ACCTACCCGCACTGTACCACTTACCTTCATCCGCGCTACCCCGACAGCGCGGCTG 5040
QY 1680 GluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisIleAsnThr 1699
DB GAGAACCGGAGACCATCATCATGACTACATCAGCTCGGAGAGATGACCAACACAG 5100
QY 1700 AlaThrAlaMetAlaGlnIleArgAlaAspMetLeuArgGlyLeuSerProArgGlySerSer 1719
DB GCCACCGCATAGCGCCAGAGAGTGAATGTGAGAGGGCCTTCGCGCGCGAGTCTCTG 5160
QY 1720 LeuAlaLeuAsnTyrAlaIleGlyProArgGlyIleIleAspLeuSerGlnValProHis 1739
DB CTGGGCACTAACTACCTGCGGCTCCCGAGGATCATGCACTGTCCCAAGTGCACAC 5220
QY 1740 LeuProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAla 1759
DB CTGGCTGTCTGTCGCGCCGACACAGGACCCGACCGCATGAGCGGCTTGGCC 5280
QY 1760 TyrLeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerPro 1779
DB TACCTCCCAACCGCGCGCGCTTCAGACGCGCACAGCAGCTCCCACTCTCCCA 5340
QY 1780 GlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGlnArgGluArgAsp 1799
DB GAGAGTCCAAACACTTGACAAACACCAACCAAGTCTCTGTCGAGGCGGAGGAGAC 5400

QY 1800 ArgAspArgGluArgAspArgAspArgGluArgLysSerIleLeuThrSerThrThr 1819
DB CGGATTCGAGAGCGGAGCCGGATCGGAGCCGGAAATCCATCTCTCACTCCACAG 5401
QY 1820 ThrValGlnHisAlaProIleThrArgProGlyThrGlnGlnSerSerGlySerSerGly 1839
DB ACGGTGAGACAGCAACCATCTGAGACTGGTACAGACAAAGCAGCGGACAGCGGC 5461
QY 1840 SerSerGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaIleGln 1859
DB AGCAGCGCGGGGTGGGGGAGCAGACAGCCCGCGCTCCCATCTCCATGCCACAG 5580
QY 1860 HisSerProIleSerProArgThrGlnAspAlaLeuGlnArgProSerValLeuHis 1879
DB CACTGCGCCATCTCCCTCGGAGCCAGAGATGCTTCACAGACAGACCCAGTGTCTTAC 5581
QY 1880 AsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeuArg 1899
DB AACACAGGATAGAGGTATCATCACGCTGTGAGCCAGCACCGCCACGGTCTGAGG 5641
QY 1900 SerThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCys 1919
DB TCCACTCCACTCTCTCAACCGTTCGCGCGCTGACATTCACCTGCGCACCACTGC 5701
QY 1920 ProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuPro 1939
DB CCACTGGGGGAGCACTCGATGGGGTCTACCTTACCTCATGAGACCCGCTTGTGCCCC 5761
QY 1940 LysGluAlaProArgValAlaArgProGluArgProAlaArgAlaAspThrGlyHisAlaPhe 1959
DB AAGAGAGCCCCCGGGTCCCGGCGAGCGGAGCCCGGAGAGACAGACCGGCAATGCTTC 5821
QY 1960 LeuAlaLysProProAlaArgSerGlyLeuGlnProAlaSerSerProSerLysGlySer 1979
DB CTGGCCAGACCCCGAGCCCGCTCCGGGCTGAGCGCCGCTCTCCCGCAGAGGCTCG 5881
QY 1980 GluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAla 1999
DB GAGCCCGGCGCTGCTGCTCTCTGTGCGACAGCCACATGCGCCGACCCCTGCG 5941
QY 2000 LysAsnLeuAlaProHisIleHisAlaSerProAspProProAlaProProAlaSerAlaSer 2019
DB ACCTACCTGCACTTCAACCGCACCGCAGCCGAGCCGCGGCGCACCTGCTGCGCTCG 6001
QY 2020 AspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuArg 2039
DB GACCCGCAACGGGAAAGCTCAAGTAACTTTTTCATCCAGGAATGGAACTCGCT 6061
QY 2040 SerLeuGlyTyrHisGlySerSerTyrSerProGlnGlyValGluProValSerProVal 2059
DB TCTTGGGTTTACAGCGAGAGCTTACAGCCCGGAAAGGGTGAAGCCGTCAGCCCTGTG 6121
QY 2060 SerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGlnGluLeuAspLys 2079
DB ACCTACCCAGTCTGACCAACAGCAAGAGGCTCCCAAGACACTGAAAGCTCGACAG 6181
QY 2080 SerHisLeuGlnGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGlyGlu 2099
DB AGCCACCTCGAGAGGGAGGTGGGCGCCAAAGCAGCAGCCCGCTGAAGCTTGGCGGAG 6241
QY 2100 AlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSerProLeu 2119
DB GCGGCCACCTCCACACTCGCGCGCGCTGTGAGAGCAGCCCTCGTCAGCCGCGTG 6301
QY 2120 LeuGlnThrAlaProGlyValLysGlyHisGlnArgValIleThrLeuAlaGlnHisIle 2139
DB CTTCAGACCGCGCCAGGGGTCAAGAGTCAACAGCGGTGTCACTTGGCCAGCACATC 6361
QY 2140 SerGluValIleThrGlnAspTyrThrArgHisIleProGlnGlnLeuSerAlaProLeu 2159
DB AGTGAAGTATACACAGAGACTACACCGGACACACCAAGCACTGACGCCACCCCTG 6421
QY 2160 ProAlaProLeuTyrSerThrPheProGlyAlaSerCysProValLeuAspLeuArgArgPro 2179


```

Db      6481  CCGGCCCCCTACTCTTCCCTGGGCGACGCTCCGCTCCGACCTCCGCGCCCA 6540
Qy      2180  ProSerApleuTyLeuProProProAspHiAgLyAlaProAlaArgLySerProHi 2199
Db      6541  CCGAGTACCTTACTCTCCGCGCCCGGACCATGTGTCCTCCGCGCTCCGCTCCAC 6600
Qy      2200  SerGluGlyGlyLeuArgSerProGluProAsnLySerValLeuGlyGlyGlu 2219
Db      6601  AGCGAGGGGCGAAGAGCTCTCCAGAGCAAAACAAGAGCTGCTGGGTGGTGAG 6660
Qy      2220  AspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHiSerArgSer 2239
Db      6661  GACGGTATTAACTGTGTCCTCCGACCGAGGGCATGACGAGCGACGACCTCCGAGT 6720
Qy      2240  AlaValThrProLeuLeuTyArgAspGlyGluGlnThrGluProSerArgMetGlySer 2259
Db      6721  GCTGTGATCCCGCTGCTGTACCGGGATGGGAGACAGAGAGCCAGACGAGATGGCTCC 6780
Qy      2260  LysSerProGlyLeuThrSerGlnProProAlaPhePheSerLyLeuThrGluSerAsn 2279
Db      6781  AAGTCTCAAGCAACACACAGCCGCGACCTCTTCAAGCAAGCTGACGAGAGACAC 6840
Qy      2280  SerAlaMetValLysSerLyLeuGlnGluIleAsnLyLeuLeuAsnThrHiAsnArg 2299
Db      6841  TCCGCGCATGTCAAGTCCAGAGCAAGATCAACAAGAGCTGACACCAACGAGCG 6900
Qy      2300  AsnGluProGluTyArgLeuIleSerGlnProGlyThrGluIlePheAsnMetProAlaIle 2319
Db      6901  AATGAGCTTAATATCAATATCAAGCAAGCTGGAGCGAGATCTTCAATATGCGCGCATC 6960
Qy      2320  ThrGlyThrGlyLeuMetThrTyArgSerGlnAlaValGlnGluHiSerThrAsn 2339
Db      6961  ACCGGAAAGAGCTTATGACTTATGAGACCAAGCGGTGACGAGAACTGCCAGACCAAC 7020
Qy      2340  MetGlyLeuGlnAlaIleIleArgLyAlaLeuMetGlyTyArgAspGlnTrpGluGlu 2359
Db      7021  ATGGGGCTGAGGCGCATATTAAGAAAGCACTCATGTGTAATATGACCAAGTGGAGAG 7080
Qy      2360  SerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAla 2379
Db      7081  TCCCGCGCTCAGCGCGCAATGCTTTTACCTCTGTAATCCAGTGCACGCTGCCGCT 7140
Qy      2380  AlaMetProIleThrAlaIleAspGlyArgSerAspHiSerThrLeuThrSerProGly 2399
Db      7141  GCTATGCCCTAACCGCTGCTGACGAGGAGTGAACAACACTCAGCTGCGAGGTGC 7200
Qy      2400  GlyGlyLeuAlaLysValSerGlyArgProSerSerArgLyAlaLysSerProAlaPro 2419
Db      7201  GCGGGGAAGGCGCAAGCTCTGGGAGACCCAGAGCGGAAAGCAAGTCCCGGCGCCG 7260
Qy      2420  GlyLeuAlaSerGlyAspArgProProSerValSerSerValHiSerGluGlyAspCys 2439
Db      7261  GGGCTGGCACTGTGGGAGCCGCGCACTCTGTCTCCCTCACTGCACTGGAGGGAGACTGC 7320
Qy      2440  AsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySer 2459
Db      7321  AACCGCGAGCGCGCTCACCAACCGGTGTGGAGAGAGCGCTCTGTCGCGAGGTTCC 7380
Qy      2460  ThrProPheProTyArgProLeuIleMetArgLeuGlnAlaGlyAlaMetAlaSerPro 2479
Db      7381  AGCGCAATCCCTCAACAACCCCTGATCATCGGCTGACGCGGGTGTCAAGGCTTCCCA 7440
Qy      2480  ProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHiSerAlaATPAsp 2499
Db      7441  CCCCCACCGGGCTCCCGCGGAGCGGCGCCCTCTGCTGCGCCCAACGCGCTGGGAC 7500
Qy      2500  GluGluProLysProLeuLeuCysSerGlnTyArgGluThrLeuSerAspSerGlu 2517
Db      7501  GAGGAGCCCAAGCACTGTCTGTGCTGCAAGTACAGACACTCTCCGACAGCGAG 7554

```

```

ID      ACN44283 standard; cDNA; 8667 BP.
XX
XX
AC      ACN44283;
XX
XX      18-NOV-2004 (first entry)
DT
XX      Human mRNA sequence hCT16428.
DE
XX      Cytostatic; carcinoma; lymphoma; cancer; human; gene; db.
XX
XX      Homo sapiens.
OS
XX      WO2003073826-A2.
PN
XX      12-SEP-2003.
PD
XX
XX      28-FEB-2003; 2003MO-US006235.
PF
XX
XX      01-MAR-2002; 2002US-00087192.
PR
XX
XX      (SAGR-) SAGRES DISCOVERY.
PA
XX
XX      Morris DW;
PI
XX
XX      WPI; 2003-328604/31.
DR
XX
XX      Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT      comprises a nucleotide sequence.
PT
PS      Claim 1; SEQ ID NO 653; 0pp; English.
XX
XX      The present invention relates to novel DNA and protein sequences which
XX      are associated with carcinomas. The sequences are useful for: (i) for
XX      screening drug candidates; (ii) for screening of bioactive agent capable
XX      of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
XX      a bioactive agent capable of modulating the activity of CAP; (iv) for
XX      evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
XX      carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
XX      carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
XX      (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
XX      determining Carcinoma Associated (CA) gene copy number. In addition, the
XX      CA genes are useful as DNA vaccines and the CAP are useful as markers of
XX      carcinoma including lymphoma. The present sequence is one such CA gene
XX      sequence. Note: This patent is an equivalent to basic Patent
XX      US2002182586A1, for which no sequence data was published
XX
SQ      Sequence 8667 BP; 1860 A; 3066 C; 2575 G; 1166 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      0      Length:      8667
Score:          13145.00      Matches:      4*      2509
Percent Similarity:      99.68%      Conservative:      1
Best Local Similarity:      99.64%      Mismatches:      7
Query Match:      99.47%      Indels:      2
DB:      11      Gaps:      0

US-09-522-753-5 (1-2517) x ACN44283 (1-8667)
Qy      1 MetSerGlySerThrGlnLeuValAlaGlnThrTyArgAlaThrGluProArgTyPro 20
Db      116 ATGTGGGATCCACACAGCTGTGGACAGACTGAGGGGCGCACTAGCCCCCTACCCG 175
Qy      21 ProHisSerLeuSerTyArgProValGlnIleAlaArgThrHisPheArgValGlyLeuLeu 40
Db      176 CCCCACAGCTTTCTTACCAAGTGCAGATGCGCCGAGCGACAGACGAGCGTGGGCTCTCG 235
Qy      41 GluTyArgIleHisSerArgAspTyArgAlaSerHisLeuSerProGlySerIleIleGln 60
Db      236 GAGTACACAGCACACTCCCGCACTATGCTCCCACTGTGCGCGGCTCCACATCCAG 235
Qy      61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyLeuGluArgSerGln 80
Db      296 CCCCAGCGGCGAGGCGCTCCCTGTGTGATTCAGACCCCGGAGATGAACGGTCCAG 355

```

QY 81 GIULENHIELEUATRGPROJUSERHISSEYTYRLEUPROGLULEUGLYYSERGIUMET 100
DB 356 GAGSTCCACCTGGGCGCAAGAGTCCCACTACCTCCGAGCTGGGAGGAGTGCAGAGATG 415
QY 101 GIUPHEIIEGLUSERYARTRARGLEUGLULEUENPROABPROLEULEUAATGPRO 120
DB 416 GAGTTTATTGAAAGCAAGGCGCTCGCTAAGAGCTCTGCTGACCCCTGCTGCACCG 475
QY 121 SERPROLEULEUALATHRGVINGINPROAAGLYSERGIUABPHEUTHRYLSAEPAGSER 140
DB 476 TAAACCTTGTGCGCCACGGGCCAGCTCTGGGAGTCTGAAGACTTCAACAGAGACCTTACG 535
QY 141 LEUTHRGILYLSLEUGLUPROVALSERPROPROSERPROPHIISTHRAAPPROGLULEU 160
DB 536 CTGACGGGCAAGCTGGAAACGGGTGTCTCCCGCCAGCCCCCGCACACTGACCTGAGCTG 595
QY 161 GIULEUVALPROBPROARGLEUSERLYSGIUGLULEUITLEINAMMETASPAAGVALAP 180
DB 596 GAGCTGTGTCGCGCACGGCTGTCCAAAGAGAGCTGATCCAGAACATGGACCGCGTGAGAC 655
QY 181 ARGGLIIEETHMETVALGIUGINGINILESERLYLEULYSYLSYSGINGINILEU 200
DB 656 CGAGAGATACCATGTGTAAGACACAGATCTTAAGCTGAAGAGAAAGACAGACAGCTG 715
QY 201 GIUGIUGIUALAALAYSPPROPROGLUPROGLULYSPROVALSERPROPROIIEGLI 220
DB 716 GAGAGAGAGGCTGCCAAGCGCGCCGAGCTGAGAACCCGCTGTACCGCGCCCATCGAG 775
QY 221 SERLYSHIARSGSERLEUVALGINILEIETRYASPGIUBAMRGLYSVALAGIUALA 240
DB 776 TCGAAGCACCGCGAGCTGTGAGATCATCTACGAGAGAACCGGAAGAAAGCTGAGAGCT 835
QY 241 ALAHISARGIILEUGIUGILEUGIYPROGLINVALGIUENUPROEUTHRYASNGINPRO 260
DB 836 GACATCGGATTTCTGAGAGGCTGGGGGCCCGAGGTGAGCTGCGCTGTACACACAGCCC 895
QY 261 SEASPTTHARSGINTYRHSIGIUBENILEYSLIASNGIUALAMETARGLYSLEU 280
DB 896 TCCGACACCGCGAGATCATGAGAACATCAAAATAAACAGGCGATGGCGAAGAAAGCTA 955
QY 281 IIELEUTHRYPHELYARARGAENHIVALAARGLYEGLINTPLYSGINLYSPHECYSGIN 300
DB 956 ACTTGTACTTCAAGGAGGAGATCAAGCTCGGAAACAATGGGAGCAGAGAGATTCTCCAG 1015
QY 301 ARGTYAASGILNEUMETGIUALALEUGIULYSYLSYVALGIUBARGIIEGIUBAENAPRO 320
DB 1016 CGCTATGACCAAGCTCATGAGGCTTGAGAGAGAAAGGTGAGCGCATCGAGAACACCCC 1075
QY 321 ARGARGARGALALYSGLUSERLYSVALAARGIUTYRTRYGIULYSGINPHEPROGIUILE 340
DB 1076 CCGCGCGCGGCGCAAGAGAGAGCAAGTGGCGGACTACTACAGAGAACGATTCCTCGAGATC 1135
QY 341 ARGLYEGLNARGGIULEUGINGIUNARGMETGINSERARGVALGIULNARGIYSGIY 360
DB 1136 CCCAACAACGCGAGCTGCAGAGAGCCATGCAAGAGAGGTGGCCACGCGGCAAGTGGG 1195
QY 361 LEUSERMETSERLIALARGSERGIUHSIGIULVALSERGIUITLEIIEASPGIYLEUSER 380
DB 1196 CTGTTCATGTGCGCGCGCGAGCGAGCAAGAGTGTCAAGATCATCGATGCTCTCTCA 1255
QY 381 GIUGINGIUBASNLEUGIULYSGINMETARGGLNLEUALAVALIIEPROMETLEUTHRY 400
DB 1256 GAGCAGAGAACTGGAAGAGAGATGCGCCAGCTGGCGCTGATCCCGCCCATGCTGTAC 1315
QY 401 ASPALAAASPGINGIUNARGIIELYSPHEIIEASMETASNGIYLEUMETALAAAPROMET 420
DB 1316 GACGCTGACAGAGCGCATCAAGTTCATCAACATGACGAGCTTATGGCGGACCCCATG 1375
QY 421 LYSVALITYLYASPARGINIVALMETASNMETTPSERGIUGINGIULYSGIUTHPHE 440
DB 1376 AAGGTGTACAAAGACCGCGAGTCAATGTGAGTGAAGAGAGAGAGAACCTTTC 1435

QY 441 ARGGIULYSYPHEMETGINHISPROLYSABENPHEGLYLEUITLEASERPHETENUGIUNARG 460
DB 1436 CGAGAGAACTTATGACAGATCCCAAGAACTTGGCTGTATGCATCATCTCTGGAGAGG 1495
QY 461 LYSTHVALAIALGICYSAVALLEUTHRYTYRTRYRLEUTHLYLSASNGIUBANTYRLYS 480
DB 1496 AMAGACGTGGCTGAGTGGCTCTCTTACTTACTTACTTACATGAAGAAATGAAACTTAAAG 1555
QY 481 SERLEUVALARGASERTYRARGARGARGIYLYSSEGIINGININGININGIN 500
DB 1556 AGCTGTGTAGACGAGAGCTATCGGCGCGCGCAAGAGCCAGAGCAACAGCAGCAG 1615
QY 501 GINGINGINGINGINGINGINGININPROMETPRAIRSERSEGTINGIUGIULYS 520
DB 1616 CAGCAGCAGACAGACAGACAGACAGCCCATGCCCCGAGCGCACGAGGAGGAGAAA 1675
QY 521 AASPGIULYSGLIULYSGLIULYSGLIULYSGLIULYSGLIULYSPPROGLIUBAN 540
DB 1676 GATGAG 1735
QY 541 AASPLYSGIUBASPLEULEULYSGLIULYSSTHRAASPTTHSERGIYGLIUBASNAEPGI 560
DB 1736 GACAGAGAGAGACTCTCTCAAGAGAGAGACAGACAGACCTCAGGGGAGACAGACGAG 1795
QY 561 LYSGLIUALAVALASERTYSGIYARGLYSTHRAASERGIINGIYARGARGLYSGIY 580
DB 1796 AAGGAGGCTGTGCGCTCCAAAGCGCGCAAAACTGTCCMAACGAGCGAGGAGAGAGCGCAAGGC 1855
QY 581 ARGIIETHRARGSERMETALASNGIUALAASNSERGIUGIUALAIEHPROGLINGIN 600
DB 1856 CGCATCACCGGCTCAATGGCTAATGAGGCCAACAGAGAGAGCATCACCCCCAGCAG 1915
QY 601 SERIAGIULEUALASERMETGIULEUBANGIUSERSEARITRPHRGIUGIUGIUMET 620
DB 1916 AGCGCGAGACTGGCTCCATGAGCTGATGAGAGTTCGCTGAGACAGAGAAAGAAATG 1975
QY 621 GIUTHPHALYLSYSGIYLEUGIUIHSIGIYARGASNTTPSERALILEALAHYMET 640
DB 1976 GAAACAGCCAAAGAAAGTCTCTGGAACAAGCGCGCACGTGGCTGCGCATCGCGCCGATG 2035
QY 641 VALIYSELYSTHRYALSERGINCYLSYSAENPHETRYPHASNTYRYSLSYARGIN 660
DB 2036 GTGGGCTCCAAAGCTGTGTGCGAGTGTAAAGCTTACTTCACTTCAAGAGAGAGCGAG 2095
QY 661 ASMLEUASPGIUILEUGINGINIHISLYSLEULYSMETGIULYSGLIUNARGASNAIARG 680
DB 2096 AACCTCGATGAGATCTTGCAGACAGACACAGCTGAGAAATGGAAGAGAGAGAACGCGCG 2155
QY 681 ARGLYSLYSLYSVALAPROVALAALASERGIUGIUALAALAPHEPROPROVALVAL 700
DB 2156 ACGAAGAGAGAAAGAACCGCGCGCGCGCCAGCGAGAGAGGCTGCATTCGCGCCGTGCTG 2215
QY 701 GIUBASPGIUGIUMETGIUALASERGIYVALISERGIYANGIUGIUGIUMETVALGIUGI 720
DB 2216 GAGAGTAGAGAGATGAGGCGTGGCGGTGAGCGGAAATGAGAGAGAGATGGTGGAGAG 2275
QY 721 ALAGIUALALEUHIISALASERGIYASNGIULVALPROARGIYGLIUCYSSEGIYPROA 740
DB 2276 GCTGAAGCTTACTATCTCTGTGGAAATGAGTGCCTCAAGAGGGAATGCATGGCCAGCC 2335
QY 741 THIVALASNASERSERASPTTHRGUSERIIEPROSERPROHISSTHRIUALAALAYS 760
DB 2336 ACTGTCAACACAGCTCAGACACCGAGACATCCCTCTCTCAACACTGAGGCGCCCAAG 2395
QY 761 AASPTTHRGIUBASNGIYPROLYSPROPROVALATHPHEGLIYALASPGIYPROPROPRO 780
DB 2396 GACACAGGCGCAGATGGGCCCAAGCCCGCACCACTTGGGCGCCACAGGCGCACCCCA 2455
QY 781 GIYTPROTHPROTHPROTHPROTHPROTHPROTHPROTHPROTHPROTHPROTHPROTH 800
DB 2456 GGGCCACCCACCCACACCGAGAGAGATCCGAGCC-CCCATGAGCCACCCCGGCTTC 2514
QY 800 RGIUALATHRGIALAPROTHPROTHPROTHPROTHPROTHPROTHPROTHPROTHPROTH 820

[illegible]

Db	3555	GCTGGCACCCCTTCACGGAGTGTAAGACAGAGACAGCTGTCCCAACGGGGCCAGGCTGGGCC	3654
QY	1180	oProGluSerLeuGlyVal ¹ ProThrAlaGlnGluValAspSerValLeuArgGly ¹ ThrAlaLe	1200
Db	3655	ACCGGAGAGCCTGGGGGTGGCCACAGCCACAGAGAGCGTCCGGTGTGAGAGGGACAGCTCT	3714
QY	1200	uGlySerVal ¹ ProGlyGlySer ¹¹ IethrLysGly ¹¹ LeProSer ¹ ThrArgVal ¹ ProSerAs	1220
Db	3715	GGGCTCACTTCCGGGCGGAGACATACCAAAAGGACATTCACAGCACACGGGTGCTCTCGGA	3774
QY	1220	pSer ¹ Ala ¹ Iethr ¹ Trp ¹ ArgGly ¹ Ser ¹¹ IethrHisGly ¹ ThrProAlaAspVal ¹¹ Leu ¹ Tyr ¹ L	1240
Db	3775	CAGCGCCATCAATACCGGGCTCTCATACCAGGACGCGCAGTGAAGTCTGTATACA	3833
QY	1240	sg ¹ Y ¹ Thr ¹ Iethr ¹ Arg ¹ Ile ¹ IleGly ¹ ValAspSer ¹ ProSer ¹ ArgLeuAsp ¹ Arg ¹ Val ¹ Arg ¹	1260
Db	3835	GGGACACCTTACACAGAGATCATCGCGAGAGACGCCAGTGTGCTTGACCCGGCCGGGGA	3894
QY	1260	uAspSerLeu ¹ ProLysGly ¹ HisVal ¹¹ Ie ¹ TyrGlyGly ¹ Val ¹ Val ¹ LeuSer ¹ Ty	1280
Db	3895	GGACAGCCTGGCCCAAGGGCCACGTCATCTAGAAGGCAAGAGGGCCACGTCCTGTCTTA	3954
QY	1280	rgLug ¹ Gly ¹ MetSerVal ¹ ThrGln ¹ GlySer ¹ Val ¹ Asp ¹ Val ¹ ArgSerSer ¹ Gly ¹ Pr	1300
Db	3955	TGAGGGGTGGCATGTCTGTGACCAGTGTCTCAAGAGAGACGGCAGAACACCTACAGACC	4014
QY	1300	oProHisGlu ¹ Thr ¹ Ala ¹ Ala ¹ Pro ¹ Lys ¹ Arg ¹ Thr ¹ TyrAspMetMetGlu ¹ Val ¹ ArgVal ¹ Ala ¹ Ar	1320
Db	4015	CCCCCATAGACGGGCGGCCCCCAAGCGCACCTTAGCATGATGAGAGGGCCGCGTGGCGAG	4074
QY	1320	gAla ¹ IleSerSer ¹ AlaSer ¹ IleGlu ¹ GlyLeuMetGly ¹ Val ¹ Arg ¹ Ala ¹ IleProProGlu ¹ Arg ¹ His	1340
Db	4075	AGCCATCTCTTCACGACGACATCGAAGGTCTCATGGCGGTGCATCCGCGGAGCGACGA	4134
QY	1340	sSerProHis ¹ Ile ¹ Leu ¹ LysGlyGlnHis ¹ His ¹ Ile ¹ ArgGly ¹ Ser ¹¹ IethrGlnGly ¹¹ Ie ¹ Pr	1360
Db	4135	CAGCCGCCACCACTTCAAGAGAGACACCATCGGGGGTCCATCACACAAAGGATCTCC	4194
QY	1360	oArgSer ¹ Trp ¹ Val ¹ Glu ¹ Val ¹ GlnGlu ¹ Asp ¹ Tyr ¹ Leu ¹ Arg ¹ Arg ¹ Glu ¹ Ala ¹ Val ¹ Leu ¹ Leu ¹ Val ¹ Ar	1380
Db	4195	TGGGTCTTACCTGAGAGGACACGAGAGACTACTCTCGTCGGAGGCCAAGCTCTTAAAGCG	4254
QY	1380	gGluGly ¹ Thr ¹ Pro ¹ Pro ¹ Pro ¹ Pro ¹ Pro ¹ Ser ¹ Arg ¹ Asp ¹ Leu ¹ Thr ¹ Glu ¹ Ala ¹ Tyr ¹ Lys ¹ Thr ¹ G	1400
Db	4255	GGAGGGACCGCTCCGCCGCCACCGCCCTTCACGGACCTGACGAGGGCTTCAAGACGCA	4314
QY	1400	nAlaLeuGly ¹ Pro ¹ Leu ¹ LysLeu ¹ Val ¹ Pro ¹ AlaHis ¹ GlyGly ¹ Leu ¹ Val ¹ Ala ¹ Thr ¹ Val ¹ Lys ¹ G	1420
Db	4315	GGCCCTGGGGCCCCCTGGAAGCTGAAGCGGGCCCAAGAGGGCTGTGTGGCCACAGGTGAAGGA	4374
QY	1420	uAlaGly ¹ ArgSer ¹¹ IethrGlu ¹ Ile ¹ Pro ¹ ArgGlu ¹ Leu ¹ Arg ¹ His ¹ Thr ¹ ProGlu ¹ Leu ¹ Asp ¹ Pr	1440
Db	4375	GGCGGGCGCGCTCCATCATGAGATCCCGCGGAGAGACTGTGGGACACAGCCCGAGCTGCC	4434
QY	1440	oLeuAla ¹ Pro ¹ Arg ¹ Pro ¹ Leu ¹ LysGlySer ¹¹ IethrGlnGly ¹ Thr ¹ Pro ¹ Leu ¹ Lys ¹ Trp ¹ As	1460
Db	4435	CTTGCGCCCGGGCGCGCTCAAGAGAGGGCTCATACCGACGGGACCCCGCTCAAGTACGA	4494
QY	1460	pThr ¹ Gly ¹ Ala ¹ Asp ¹ Thr ¹ Thr ¹ GlySer ¹ Lys ¹ Val ¹ His ¹ Asp ¹ Val ¹ ArgSer ¹ Leu ¹ Ile ¹ GlySer ¹ Pr	1480
Db	4495	CACCGGCGCGCTCACACTGGCTCCAAAAAGACAGATGAGCTCTCTCATCGGCACACCC	4554
QY	1480	og ¹ Lys ¹ Arg ¹ Thr ¹ Pro ¹ Pro ¹ Val ¹ His ¹ Pro ¹ uAsp ¹ Val ¹ Met ¹ AlaAsp ¹ Ala ¹ Arg ¹ Ala ¹ Glu ¹	1500
Db	4555	CGGCGGACGTTCCACCCGTCGACCCGCTGAGATGTATGGCCACACCCGGGACATGGA	4614
QY	1500	uArg ¹ Ala ¹ Cys ¹ Trp ¹ Glu ¹ Leu ¹ Ser ¹ Leu ¹ Lys ¹ Arg ¹ Pro ¹ Gly ¹ Thr ¹ Ala ¹ Asp ¹ Ser ¹ Ser ¹ Gly ¹	1520
Db	4615	ACGTCCTCTGTAAGAGAGAGCGTGAAGACCGGCCAGGACCGCCAGACGCTCGGGGGG	4674
QY	1520	ySer ¹¹ Ile ¹ Ala ¹ Arg ¹ Gly ¹ Ala ¹ ProVal ¹¹ Ile ¹ Val ¹ ProGlu ¹ Leu ¹ Gly ¹ Lys ¹ Pro ¹ Arg ¹ Leu ¹ Ser ¹ Pr	1540
Db	4675	CTTCATTTGCGGCGCGGCCCGCGGTCAATTGTGGCTGGCAAGCGCGGCGAGACCC	4734

QY 1540 oLeuThrTYrgLuAspHisgLYaLaProPhealagLYhiSLeuProArgLYseProVa 1560
Db 4735 CCTAACCTATGAGGACCAAGGGGACCCCTTGGCGGCACTCCACAGAGTTGCGCCG 4794
QY 1560 lThrMetArgGLuProThrProArgLeuGlnGluGlySerLeuSerSerSeLySaLaSe 1580
Db 4795 GACCAAGCGGGAGCCCAAGCCGCGCTGCAAGAGGAGGCTTTCGTCCAGCAAGGCATC 4854
QY 1580 rGlnAspArgLYLeuThrSerThrProArgLuIlLeaLYseSerProHisSerThrya 1600
Db 4855 CCAAGACCCAAAGCTACGTCGACGCTCTGTAGATGCGCAAGTCCCGCAGACCCGT 4914
QY 1600 lProGluHisSLeuProHisProHisSerProTYrgLuHisSLeuLeuArgLYVaLYserG 1620
Db 4915 GCCCGAGCAACACCAACCCCATCTCGCCCTATGAGCACTGCTTCGGGGCGTAGTG 4974
QY 1620 yValaLpLeuTYrArgSerHisSLeuProLeuLaPheAspProThSerLLeProArg 1640
Db 4975 GGTGACCTGTATCGAGCCACATCCCTCGGCTTCGACCCCACTCCATACCCCGCG 5034
QY 1640 yLLeProLeuAspAlaLaLaLaTYrTYrLeuProArgHisSLeuAlaProAspProTh 1660
Db 5035 CATCCCTCTGAGAGCAGCCGCTGCTATACCTGCCCCAGACCTGGCCCCCAACCCAC 5094
QY 1660 rTYrProHisSLeuTYrProProTYrLeuIlLeaArgLYrTYrProAspThraLaLeuG 1680
Db 5095 CTACCCGCACTGTACCCACCTACCTCATCCGCGCTACCCGAGCAAGCGCGCTGGA 5154
QY 1680 uAsnArgGLnThrLLeaAsnAspTYrLLeThrSerGlnGlnMetHisSAsnThraL 1700
Db 5155 GAACCGGCAAGCATCATCATATGATCATACCTGCGAGCATGACACCAACCGCG 5214
QY 1700 aThraLaMetAlaGlnArgAlaAspMetLeuArgLYLeuSerProArgGlnSerSe 1720
Db 5215 CACCGCAATGGCCCAAGCACTGATATGCTGAGGGGCTTCGCCCGCGAGTCTCGCT 5274
QY 1720 uAlaLeuAsnTYrAlaAlaGLYProArgLYLeIlLeaAspLeuSerGlnVaLProHis 1740
Db 5275 GGACATCAACTACGTGGGGGTCCCGAGGCATCATGACCTGTCCCAAGTCCACACCT 5334
QY 1740 uProValLeuValProProThrProGLYThrProAlaThraLaMetAspArgLeuAlaTY 1760
Db 5335 GCGTGTGCTGTGTCGCCCGCAACAGGCAACCCAGCAACCGCATGAGCGCTTCCTCA 5394
QY 1760 rLeuProThraLaProGlnProPheSerSerArgHisSerSerSerProLeuSerProG 1780
Db 5395 CTTCCCAACCGGCCCCAGCCCTTCAGAGCCGCCACAGAGCTCCCACTCTCCCAAG 5454
QY 1780 yGlyProThrHisSLeuThryAspProThrThrThrSerSerSerGluArgGLuArgAsp 1800
Db 5455 AGGTCCAAACACTTACAAACCAACCAACCAACGTCGTGTCGAGCGGAGGAGCAACCG 5514
QY 1800 gAspArgGLuArgAspArgAspArgGLuArgGLuYrSerLLeuThrSerThrThr 1820
Db 5515 GAATTCAGAGCGGAGCGGAGTGGAGCGGAAAGTCCATCTCACTCCACCGCAACG 5574
QY 1820 rValGluHisLaProIlLeThraPArgProGlyThrGlnGlnSerSerGlySerSerGlySe 1840
Db 5575 GGTGAGCAAGCAACCATCTGAGACCTGTACAGAGCAAGAGCGGCGAGCGGCGAG 5634
QY 1840 rSerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisSAlaHisgLnH 1860
Db 5635 CAGCGCGGGGGGTGGGGGAGCAGCAGCGCGCTCCCACTCCCATCCACAGCA 5694
QY 1860 sSerProLLeSerProArgThrGlnAspAlaLeuGlnGlnArgProSerVaLLeuHis 1880
Db 5695 CTGCGCCCATCTCCCTCGAGCCAGAGATGCTTCAGCAGAGAACCAAGTGTCTTCA 5754
QY 1880 nThrGlyMetLYsGLYLeIlLeThraLaValGluPProSerLYrProThrVaLLeuArgSe 1900
Db 5755 CACAGGCAATGAAGGATATCATCACCGCTGTGAGGCCAGCAGCCCAAGGCTCTGAGG 5814

QY 1900 rThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCySer 1920
Db 5815 CACTTCACCTCTCTCAACCCGTTGCGCCGGCTGCACATTCACACCTGCACCAACCGCC 5874
QY 1920 oLeuGlyGlyThrThrLeuAspGLYVaLYrProThrThrMetGluProValLeuLeuProLY 1940
Db 5875 ACTGGGGGACCTCTGATGAGGATCTACCTTACCTCATGAGACCCGCTTGTGCTGCCAA 5934
QY 1940 gLuLaProArgVaLaLaArgProGluArgProArgAlaAspThrGlyHisSAlaPheLe 1960
Db 5935 GAGAGGCCCGGGGTGGCGGCGAGAGGGCCGAGAGACACCGGCGCATGCTTCCT 5994
QY 1960 uAlaLYrProProAlaArgSerGlyLeuGluProAlaSerSerProSerLYrGlyserG 1980
Db 5995 CCGCAAGCCCGCAAGCCGCTCGGAGTGAAGCCGCTCTCCCGCCAGCAAGGCTCGGA 6054
QY 1980 uProArgProLeuValProProValSerGLYHisSAlaThrLLeaArgThProAlaTY 2000
Db 6055 GCCCGGCCCCCTTAGTCTCTCTGTCTGAGCAAGCCACATGCGCGCACCTTCGGA 6114
QY 2000 sAsnLeuAlaProHisSHisSAlaSerProAspProProAlaProProAlaSerAlaSerAs 2020
Db 6115 GAACCTCGCACCTCACCAAGCAGCCAGCCGCGAGCCGCGCGCACCTCGGCTCGGA 6174
QY 2020 pProHisSArgGLuYrThrGlnSerLYrProPheSerLLeGlnGluLeuGluLeuArgSe 2040
Db 6175 CCGGCAACCGGAAAGACTCAAGTAAACCTTTTCATCCAGGAATCGAAATCCGTTTC 6234
QY 2040 rLeuGlyTYrHisSGLYSerSerTYrSerProGluGlyVaLYGluProValSerProValSe 2060
Db 6235 TCTGGGTTTCCACGGGAGAGCTACAGCCCGAAAGGAGTGAAGCCGTGACCTGTAG 6294
QY 2060 rSerProSerLeuThrHisSAspLYsGLYLeuProLYrHisSLeuGluGluLeuAspLYSe 2080
Db 6295 CTACACCACTGACCCCAAGCAAGAGGGCTCCCAAGCACTGGAAGAGACTGCACAAGAG 6354
QY 2080 rHisSLeuGluGlyGluLeuArgProLYsGlnProGLYrProValLYsLeuGlyGlyGlu 2100
Db 6355 CCACTGTGAGGGGAGCTCGGCGCCCAAGCAGCGCCCGGTGAAGCTTGGGGGAGGCG 6414
QY 2100 sAlaHisSLeuProHisSLeuArgProLeuPProGluSerGlnProSerSerSerProLeu 2120
Db 6415 CGCCCACTCCCAACCTGCGGCGCTGCTGAGAGCCAGCCCTCATCCAGCCCGCTGCT 6474
QY 2120 uGlnThraLaProGLYVaLYsGLYHisSgLnArgVaLYrThrLeuAlaGlnHisSLeSe 2140
Db 6475 CCAAGCGGCCCAAGGGGTCAAAAGTCAACAGGGGGGTGCACCTTGCCAGACATTCAG 6534
QY 2140 rGluVaLYLeThrGlnAspTYrThrArgHisSHisSProGlnGlnLeuSerAlaProLeu 2160
Db 6535 TGAAGTCATCACACAGGACTACACCCGCGCAACCAACCAAGCACTCAGCGCAACCTTGCC 6594
QY 2160 oAlaProLeuTYrSerPheProGLYAlaSerCySPProValLeuAspLeuArgArgPro 2180
Db 6595 CGCCCCCTCTACTCTCTCTGAGGCGAGCTGCCCCGTTCTGAGCTTCGCCGCCAC 6654
QY 2180 oSerAspLeuTYrLeuProProProAspHisgLYaLProAlaArgLYrSerProHisSe 2200
Db 6655 CAGTAGCCTCTACTCTCCGCCCGCCGAGCAATGATGCCCCGGCGGTCTCCGCCACAG 6714
QY 2200 rGluGlyGlyLYrArgSerProGluProAsnLYrThrSerValLeuGlyGlyGlyGlu 2220
Db 6715 CGAAGGGGCAAGAGGTCTCCAGAGCAAAAGCTGTGGCTTGGGTGGTGGTAGGA 6774
QY 2220 pGlyLLeGluProValSerProProGluGluYrMetThrGluProGlyHisSerArgSerAl 2240
Db 6775 CGGTATGAACCTGTGTCCTCCACCGAGGGCAATGAGCGAGCCAGGGCACTCCGGAGTGC 6834
QY 2240 aValTYrProLeuLeuTYrArgAspGLYgLuGlnThrGluProSerArgMetGlySerLY 2260
Db 6835 TGTGTACCCGCTGCTGTACCGGGATGGGGAACAGAGCCCAAGCAGATGGGCTCCAA 6894
QY 2260 sSerProGlyAsnThrSerGlnProProAlaPhePheSerLYrLeuThrGluSerAsnSe 2280

```

Db      6895 GTCTCCAGGCAACCAAGCCAGCCGACCTTCTTGAAGCAAGCCAGCAACATC 6954
Qy      2280 TALAmetVallyserlyserlysgngllyleantlylyseunenthrthsaenaygas 2300
Db      6955 CCGCATGCTCAAGTCCAGAAAGCAAGATCAACAGAGGTGAACCCCAACCCGAA 7014
Qy      2300 ngiuproglytyrbanileserjnpoglythrghuliepheanmetproalaileth 2320
Db      7015 TGAGCCTGAATACAAATATACAGCCGCTGGAGCGAGATCTTCAATATGCCGATCAC 7074
Qy      2320 rgllythglyleumethrtyrargserglnalvalinglnghisalasertthraenme 2340
Db      7075 CGGAACGCGCTTATGACCTATAGAACCGAGCGCTCAGGAACATCCAGACCAACAT 7134
Qy      2340 tgllyleuglnalalilelearglysalaleumetglylyethrargpintpugluse 2360
Db      7135 GGGGCTTGAGGCGCAATATAGAAAGGCACTCATGGTAAATATACAGATGGGAAGATC 7194
Qy      2360 rproproleuseralaaanlaapheanproleuasnlaaserlaserleuproalial 2380
Db      7195 CCGCGCGCTCAGCGCCAAATGCTTTAACTCTGAAATGCCAGTCCAGCCCGCTGC 7254
Qy      2380 aMetProIleThrAlaAlaapglyargserarphisthrleuthrserproglylgl 2400
Db      7255 TATGCCCATAAACGCTGCTGACGAGCGAGGAGACACACTCCTCCGACAGGTGCGG 7314
Qy      2400 ygllyvalalylvalserglyargproseraserarglyvalalaserproalaprogl 2420
Db      7315 CGGAAGGCGCAAGGTCTCTGACAGACCCAGACGCCAAAGCCAAAGTCCCGGCGCGG 7374
Qy      2420 yleualaserglyasparproproservalisersevalhisergluglyasparCyas 2440
Db      7375 CTGGGATCTGGGGACCGGCCACCTCTGTCTCTCTCATGTCACTCGAGGAGACTCAA 7434
Qy      2440 natgahgthproleuthraanargvalitpgluasparproseraserlajlserth 2460
Db      7435 CCCCAGGACCGCCCTCAACCAACCGCGTGGAGAGACAGGCCCTCGCCAGGTTCCAC 7494
Qy      2460 rpropherothyraenproleuilemetargleuglnalglyvalmetlaserpropr 2480
Db      7495 GCCATTCCCTTACAAACCCCTGATCATGCGGACGCGGCTGTCATGCTTCCACC 7554
Qy      2480 opProPoglyleuProalaglyserglyproleualaglyprohibihialatTPAapgl 2500
Db      7555 CCAACCGGCGCTCCCGCGGCGAGCGGGCCCTCGCTGGCCGCCACCAACGCTGGACGA 7614
Qy      2500 ugiuprolypProleuencyserserglnyrgiuthrleuseraspserglu 2517
Db      7615 GGAGCCCAAGCACTGCTGTGCTGCGAGTACAGACACTCTCCGACAGCGAG 7666

```

RESULT 6

ADQ84524 standard; cDNA; 8533 BP.

ADQ84524;

07-OCT-2004 (first entry)

Human tumour-associated antigenic target (TAT) cDNA sequence #1338.

human; tumour-associated antigenic target; TAT; cytostatic; gene therapy;

cancer; cell proliferative disorder; gene; ss.

Homo sapiens.

WO2004060270-A2.

22-JUL-2004.

15-OCT-2003; 2003WO-US029126.

18-OCT-2002; 2002US-0418988P.

XX (GETH) GENENTECH INC.
PA (WUTD/) WU T D.
PA (ZHOU/) ZHOU Y.
XX Wu TD, Zhou Y,
XX WPI; 2004-534300/51.
XX
PT New nucleic acid molecule and encoded polypeptide, for diagnosing,
PT preventing or treating cell proliferative disorders such as cancer.
XX
PS Claim 1; SEQ ID NO 1338; 5504pp; English.

CC The present invention describes an isolated tumour-associated antigenic
CC target (TAT) nucleic acid comprising: (a) any of 4622 nucleotide
CC sequences (see SEQ ID NO:1 to 4622); (b) the full-length coding region of
CC (a); (c) the complement of (a) or (b); (d) a sequence that has 80%
CC sequence identity to (a)-(c); or (e) a sequence that hybridises to (a)-
CC (c). Also described: (1) an expression vector comprising the above
CC nucleic acid; (2) a host cell comprising the above expression vector; (3)
CC a process for producing a polypeptide; (4) an isolated polypeptide
CC comprising: (a) an amino acid sequence encoded by any of the above
CC nucleotide sequences; (b) an amino acid sequence encoded by the full-
CC length coding region of the above nucleotide sequences; or (c) a sequence
CC having at least 80% identical to (a) or (b); (5) a chimeric polypeptide
CC comprising the above polypeptide fused to a heterologous polypeptide; (6)
CC an isolated antibody that binds to the above polypeptide; (7) a process
CC for producing the antibody; (8) an isolated oligopeptide that binds to
CC the above polypeptide; (9) a tumour-associated antigenic target (TAT)
CC binding organic molecule that binds to the above polypeptide; (10) a
CC composition of matter comprising the above (chimeric) polypeptide,
CC antibody, oligopeptide or TAT binding organic molecule, in combination
CC with a carrier; (11) an article of manufacture comprising a container and
CC the composition of matter contained within the container; (12) methods of
CC inhibiting the growth of a cell that expresses the above protein, where
CC the growth of the cell is at least in part dependent upon a growth
CC potentiating effect of the above protein; (13) a method of
CC therapeutically treating a mammal having a cancerous tumour comprising
CC cells that express the above protein; (14) a method of determining the
CC presence of a protein in a sample suspected of containing the protein
CC described above; (15) methods of diagnosing the presence of a tumour in a
CC mammal; (16) a method for treating or preventing a cell proliferative
CC disorder associated with increased expression or activity of the above
CC protein; and (17) a method of binding an antibody, oligopeptide or
CC organic molecule to a cell that expresses the protein described above.
CC The TAT sequences have cytostatic activities, and can be used in gene
CC therapy. The composition and methods are useful for diagnosing,
CC preventing or treating cancer. The composition is also used for preparing
CC a medicament for the therapeutic treatment or diagnostic detection of a
CC cell proliferative disorder or cancer. The present sequence represents a
CC human TAT cDNA sequence from the present invention.

XX Sequence 8533 BP; 1836 A; 3034 C; 2528 G; 1135 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0 Length: 8533
Score: 13119.50 Matches: 2506
Percent Similarity: 99.56% Conservative: 1
Best Local Similarity: 99.52% Mismatches: 7
Query Match: 99.28% Indels: 5
DB: 13 Gaps: 1

US-09-522-753-5 (1-2517) X ADQ84524 (1-8533)

```

Qy      1 MetSerGlySerThrGlnLeuValAlaGlnThrTTPArgAlaThrGluProArgTyrPro 20
Db      2 ATGTGCGGATCCACACAGCGCTGTGGACAGACGTGAGGGCCACTAGGCCCGCTACCGG 61
Qy      21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrProValGlyLeuLeu 40
Db      62 CCCACAGCGCTTCTTCAACCAAGTGCAGATGCGCCGAGACGACAGGACGTGGGCTCTCTG 121

```

QY 41 GIUTYRGlnHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
DB 122 GAGTACACAGACACACTCCCGGACATATGCTCCCACTCTCGCGCTCCATCATCCAG 181
QY 61 ProGlnArgArgProSerIleuSerGlnPheGlnProGlyAsnGlnArgSerGln 80
DB 182 CCCCAGCGGAGGAGCCCTCCCTGCTGTGTGAGTTCCAGCCCGGAAATAGACGCTCCAG 241
QY 81 GIULeuHisLeuArgProGlySerHisSerTyrLeuProGlyIleuGlyIleSerGlnMet 100
DB 242 GAGCTCCACCTGGCGGACAGTCCCACTATACCTCCGAGTGGGAAAGTAGAGAG 301
QY 101 GIUpeHleGlySerIleArgProArgLeuGlyIleuLeuProAspProLeuLeuArgPro 120
DB 302 GAGTTCATTTGAAAGCAAGCGCCCTCGGCTAGAGCTCTGCTGACCCCTGCTGCGACCG 361
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspIleuThrIleAspArgSer 140
DB 362 TCACCCCTGCTGGCCACGGGCCAGCTGGGGATCTGAAGACTCACCAAGACCGTATGC 421
QY 141 LeuThrGlyIleLeuGlyIleProValSerProProSerProProHisThrAspProGlyLeu 160
DB 422 CTGACCGGAGAGTGGAAACCGGTGTCTCCCGGACGCCCGGACACTGACCTGAGCTG 481
QY 161 GIULeuValProProArgLeuSerIleArgGlyIleuIleGlnAsnMetAspArgValAsp 180
DB 482 GAGCTGTGTCGGCCACAGCGCTGTCCAGAGAGAGCTATCCAGAACATGACCCGCTGAGAC 541
QY 181 ArgGlyIleThrMetValGlyGlnGlnIleSerIleLeuIleIleValSerGlnGlnGln 200
DB 542 CAGAGATACCACTGATGAGCAGCAGATCTTAACCTGAAGAAAGACAGCAACGCTG 601
QY 201 GIUGlyGlyAlaAlaIleArgProProGlyIleProGlyIleProValSerProProIleGln 220
DB 602 GAGAGAGAGGCTGCCAAGCCGCCGAGCTCGAAGAACCCGCTGTCAACGCCGCCATCGAG 661
QY 221 SerIleHisArgSerLeuValGlnIleIleTyrAspGluAsnArgIleValAspIle 240
DB 662 TCGAAGCACCGCAGCTGTGTCAGATCATCTACGACGAAACCGGAAGAAAGCTGAAAGCT 721
QY 241 AlaHisArgIleLeuGlnGlyLeuGlyIleProGlnValGlyLeuProIleuThrAsnGlnPro 260
DB 722 GACATCGGATTCGGAAAGCCTGGGGCCCGGAGTGAAGTCCGCTGTACCAACGAGCC 781
QY 261 SerAspThrArgGlnTyrHisGlnAsnIleIleValIleAsnGlnAlaMetArgIleValLeu 280
DB 782 TCCGACACCCCGGAGATCATGAGAACATCAAAATTAACAGGCGATGCGGAAGAGCTA 841
QY 281 IleLeuTyrPheIleArgArgAsnHisAlaArgIleGlnTyrIleGlnIlePheCysGln 300
DB 842 ATCTTGTACTTCAAGAGAGGAATCAAGCTCGGAAACAATGGAGACGAAAGTTCTCCAG 901
QY 301 ArgTyrArgGlnIleuMetGlnAlaLeuGlyIleValGlyValGlyIleGlnAsnAsnPro 320
DB 902 CGCTATGACAGCTCATGAGGCTCTGGAGAAAGAGTGAAGCATCGAAGAACACCC 961
QY 321 ArgArgArgAlaIleArgIleSerIleValArgGlyTyrTyrGlyIleGlnPheProGlyIle 340
DB 962 CCGCGCGCGGCCAAGAGAGCAAGGTCGCGAGTACTACAGAAAGCAAGTTCTCTGAGATC 1021
QY 341 ArgIleGlnArgGlyIleuGlnGlnArgMetGlnSerArgValGlyIleGlnArgIleSerGly 360
DB 1022 CCCAAGCACCGCAGCTGAGAGGCAATGCAAGACAGAGGTGGGCCAGCGCGCACTGGG 1081
QY 361 LeuSerMetSerIleAlaArgSerGlyIleGlyValSerGlnIleIleAspGlyLeuSer 380
DB 1082 CTGTTCATGTGGCGCCCGCAGCAGACAGAGGTGTCAAGATCAATGATGCTCTCA 1141
QY 381 GIUGlnGlyAsnLeuGlyIleArgGlnMetArgGlnLeuAlaValIleProPheMetLeuTyr 400
DB 1142 GAGCAGAGAAACTGAGAAAGAGATGCGCGAGCTGGCGGTGATCCCGCCAGCTGTAC 1201
QY 401 AspAlaAspGlnGlnArgIleIleValSerPheIleAsnMetAsnGlyLeuMetAlaAspPheMet 420

DB 1202 GACGCTGACACAGACCGCATCAAGTTCAATCAACATGAAGGGCTTATGGCCACCCACAG 1261
QY 421 LysValIleTyrLysAspArgGlnValMetAsnMetTPSerGlyGlnGlyIleValPhe 440
DB 1262 AAGGTGTACAAAGACCGCAGGTCTATGAACATGTGAGTGAAGACAGAGAAAGAGACCTTC 1321
QY 441 ArgGlyIlePheMetGlnHisProIleAsnPheGlyLeuIleAlaSerPheLeuGlyArg 460
DB 1322 CCGGAGAAAGTTATCAGCATCCCAAGAACTTTGGCTGTGATGCAATCTCTCGGAGAGG 1381
QY 461 LysThrValAlaGlyCysValLeuTyrTyrTyrIleuThrIleValAsnGlyIleAsnTyrLys 480
DB 1382 AAGACAGTGGCTGAGTGGCTCTCTATTATCTACCTGACTAAAGAAATAGAACTATTAAG 1441
QY 481 SerLeuValArgArgSerTyrArgArgArgGlyIleSerGlnGlnGlnGlnGlnGln 500
DB 1442 AGCTGTGTAGACGAGAGCTATGGCGCCCGGCAAGGACAGCAGACAGACAGACAGAG 1501
QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
DB 1502 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1561
QY 521 AspGlyIleValGlyGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGly 540
DB 1562 GATGAGAAAGAGAGAAAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1621
QY 541 AspLysGlyAspLeuLeuIleValSerIleThrAspAspThrSerGlyIleValAsnAspGly 560
DB 1622 GACAAAGAAAGACCTCTCTCAAGAGAAAGACAGACACTCAGGAGGAGAGACAAACAGAG 1681
QY 561 LysGlyAlaValAlaSerIleGlyIleArgIleThrAlaAsnSerGlnIleValArgIleGly 580
DB 1682 AAGAGAGCTGTGGCTCCAAAGCCCGCAAAATGTGCCAACAGCAGAGAGAGAGAGAGAG 1741
QY 581 ArgIleThrArgSerMetAlaAsnGlnAlaAsnSerGlyIleValIleThrProGlnGln 600
DB 1742 CGCATCACCGGCTCAATGCTATATGAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1801
QY 601 SerAlaGlyLeuAlaSerMetGlyLeuAsnGlySerSerArgTyrThrGlyGlyIleMet 620
DB 1802 AGCCCGAGCTGGCTCCATGAGCTGAAATGAGATGATTCGCTGAGCAGAAAGAAATG 1861
QY 621 GIUThrAlaIleLysGlyLeuLeuGlnHisGlyIleArgMetTPSerAlaIleAlaArgMet 640
DB 1862 GAAACAGCCAAAGAAAGTCTCTGGAACAAGCCCGCAATGTGTGGCTCATCGCCCGAGAG 1921
QY 641 ValGlySerIleThrValSerGlnCysIleAsnPheTyrPheAsnTyrIleValArgGln 660
DB 1922 GTGGGCTCCAAAGACTGTGTGAGGTGAAGAACTTCACTTCAACTACAAAGAGAGAG 1981
QY 661 AsnLeuAspGlyIleLeuGlnGlnHisIleLysLeuLysMetGlyIleGlyIleValArgAsnAlaArg 680
DB 1982 AACCTCGATGAGATCTTTCAGCAGACACAGCTGGAAGTGAAGAAAGAGAAAGAGCGCGG 2041
QY 2042 ACGAAGAAAGAAAGAGCGCGCGCGCGCGCGCGAGAGAGAGCTGCAATTCGCGCCGCTGTG 2101
QY 701 GIUAspGlyIleMetGlyAlaSerGlyValSerGlyIleGlyIleGlyIleGlyIleGlyIle 720
DB 2102 GAGGATGAGAGATGAGAGCGTGGAGCTGAGCGGAAAGAGAGAGAGAGAGAGAGAGAG 2161
QY 721 AlaGlyAlaLeuHisAlaSerGlyIleGlnGlyValProArgGlyIleCysSerGlyProAla 740
DB 2162 GCTGAAGCTTCAATCCCTCTGGAAATGAGTGCACAGAGGGAATGCAATGGCCCAAGCC 2221
QY 741 ThrValAsnAsnSerSerAspThrGlySerIleProSerProHisThrGlyAlaAlaIleVal 760
DB 2222 ACTGTCAACAACAGCTCAAGACACGAGAGACATCCCTCTCTCACTAGAGGCGCCCAAG 2281
QY 761 AspThrGlyGlnAsnGlyProIleProAlaThrLeuGlyValAspGlyProProPro 780

Db	2282	GACACAGGGCAGAAATGGGGCCCAAGCCCCAGCCACCTTGGGGCGGCAGAGGGCCACCCCA	2341
Qy	761	GIYPRProThrProPro-ArgArgThrSerArgAlaProIleGIuProThrProIaSe	800
Db	2342	GGGCGACCAACCCCAACCAACCGGAGGACATCCGGCC-CCACTGAGCCACCCCGGCTTC	2400
Qy	800	IGLUALaThrGIYAlaProThrProProProAlaProProSerProSerAlaProProPr	820
Db	2401	TGAAACCAACCGGAGCCCTTACGCCCCCAACAGCACCCCATCGCTTCGACCTCC	2460
Qy	820	oValValProLybsGIuGIuLybsGIuGIuThrAlaAlaProProValGIuGIuGI	840
Db	2461	TGTGATCCCAAGAGAGAGAGAGAGAGAGACCGCAGAGGGCCCCAGTGGAGAGAGG	2520
Qy	840	YGIuGIuGIuLybsProProAlaAlaGIuGIuLeuValAlaIAspThrGIYLybAlaGIuGI	860
Db	2521	GGAGGAGCAGAAAGCCCCCGGGCGCTGAGAGAGCTGGCACTGGACACAGGAAAGCCGAG	2580
Qy	860	uProValysSerGIuCYrThrGIuGIuAlaGIuGIuGIuProAlaLybGIYLybAspAl	880
Db	2581	GCCCCCTCAAGAGCGAGTGCACCGAGAGAGCCGAGAGGGGGCGCGCAAGGCGCAAGACG	2640
Qy	880	aGIuAlaIaGIuAlaIaThrAlaGIuGIuAlaIaLeuLybAlaGIuLybLybsGIuGIYLyse	900
Db	2641	GGAGGCGCCTGAGGCGCACGCGCGAGGGGGCGCTCAAGGCAAGAAAGAGGGCGGGAG	2700
Qy	900	IGLYArgAlaThrThrThralysSerSerGIYAlaProIleAspSerAspSerSerAlaTh	920
Db	2701	CGGCAAGGCGCCACACAGCCAGAGAGCTCGGGCGCCCCCAAGACAGCACTCAAGTCTAC	2760
Qy	920	TCysSerAlaAspGIuValaAspGIuAlaGIuGIuYGIYAspLybAspArgLeuLeuSerPr	940
Db	2761	CTGCGAGTCAACAGCGAGTGGATGAGGCGCGAGGGCGGCGACAAAGAACCGGCTGTCTCCC	2820
Qy	940	oArgProSerLeuLeuThrProThrGIYAspProArgAlaAspAlaSerProGIuLybPr	960
Db	2821	AAGCGCCAGCCTCTCAACCCGAGCTGGGGAGCCCCCGGGCCAAATGCCTCACCCCAAGAGC	2880
Qy	960	OLeuaAspLeuLybsGIuLeuLybsGIuArgAlaAlaAlaIleProProIleGIuValaThrLy	980
Db	2881	ACTGGAAGCTGAAGCAGAGCTGAAGACAGCAGCGGCTGCAATCCCCCATCAAGTCAACCA	2940
Qy	980	sValIhIGIuProProArgGIuAspAlaAlaProThrLybProAlaProProAlaProPr	1000
Db	2941	AGTCATAGACCCCCCGGAGAGAGAGCGAGCTCCACCAAGCAGCACTCCCCAGCCCCAC	3000
Qy	1000	oProProGIuAspLeuGIuInProGIuSerAspAlaProGIuInProGIYSerSerProAr	1020
Db	3001	GCCACCGCAAAACCTGCAGCGCGAGAGAGAGCGCCCTCAAGAGGCTGGCAGACACCCCCG	3060
Qy	1020	GGLYLybsSerArgSerProAlaProProAlaAspLybGIuAlaIaIaGIuAlaGI	1040
Db	3061	GGGCAAGAGCAGAGACCCCGGACCCCCCGCGCAAGAGGCTTCGAGCGGAGGCCA	3120
Qy	1040	nLYsLeuProGIYAspProProCYrTTPThrSerGIYLeuProIbProValaProProAr	1060
Db	3121	GAACTGCTGGGGAGCCCCCTTGCTGGACTTCGGGCTGGCTTCCTCCCGAGCCCCCG	3180
Qy	1060	gGIuValaIleLybAlaSerProIhAlaProAspProSerAlaIAspSerTYAlaProPr	1080
Db	3181	TGAGGTGATCAAGGAGCTCCCGGCAATGCCCGGAGCCCTCAAGCTTCCTCAAGCTCCAC	3240
Qy	1080	oGLYhIAspLeuProLeuGIYLeuIhIAspThrAlaIAspProValaLeuProAspProPr	1100
Db	3241	TGGTCAACCACTGGCCCCCTGGGCTTCATGACACTGCCCCGGCGCTCTGGCGGCGCCAC	3300
Qy	1100	oThrIleSerAspProProProLeuIleSerSerAlaLybAlaIAspProValaLeuGIuAr	1120
Db	3301	CACCATCTCCAAACCGGCTCCCTCATCTCTTGCCAAAGACCCCAAGCGTCTCGAGAG	3360
Qy	1120	gGIuIleGIYAlaIleSerGIuGIYmeSerValGIuLeuIhIAspProTYSerGIuIhI	1140
Db	3361	GCAATATAGTGCATCTCCCAAGAAATGTGGTCAAGTCCACCTCCCGTACTCAGAGCA	3420

QY	1140	gAlaAlaValaAaProValGlyProValaThhMeGlyLeuPProLeuPProMetAaPProValyLy	1160
Db	3421	TGCAAGAGCCCGGGGGGGCCCTGTCCATGATGAGGGGTGCCCTGTCCATGAGACCCCAAAAA	3480
QY	1160	sLeuAlaPProPheSerGlyValaLySGInGlyGlnLeuSerProArgGlyGlnAlaGlyPr	1180
Db	3481	CCTGGACACCTTTCACCGGAGAGTGAAGACGAGACACTGTCCCAACGGGGCCAGGCTGGCC	3540
QY	1180	oProGlnSerLeuGlyValaProThraGlnGlnAlaSerValLeuArgGlyThraAlaLe	1200
Db	3541	ACCGAGAGAGCTGGGGGGTGGCCCAAGCCCAAGAGAGGGCGTCCGTGTCGAGAGGACAGCTCT	3600
QY	1200	uGlySerAlaPProGlyGlySerIleThhLyGlyIlePProSerThraGlyProSerAs	1220
Db	3601	GGGCTCAATTCGGAGCGGAGACATCACCAAGGCAATTCACGACACAGGGAGGCGCTCGGA	3660
QY	1220	pSerAlaIleThhThraArgGlySerIleThhAsGlyThhProAlaAaPValaLeuTyLy	1240
Db	3661	CAGCGCCATCATACCGGGCTCGTCATCATCCCAAGGACCGCATGTACGTCCTGTACAA	3720
QY	1240	gGlyThhIleThhArgIleIleGlyGlnAaPProSerProSerArgLeuAaPArgGlyArgGly	1260
Db	3721	GGGACCATCATCCAGAGTATCATCGGAGGAGACAGCGGCAATGCTTTGACCGCGGCGGGA	3780
QY	1260	uAaPProLeuPProGlyGlyThhValaIleTyGlyGlyLybLyGlyThhValaLeuSerTy	1280
Db	3781	GGACAGCTGCGCCCAAGGGCCACGTCATCTAGAAAGCAAGAAAGGCCACGCTTGTCTTA	3840
QY	1280	rGlyGlyGlyMeSerValThhGlnCybSerLyGlnAaPArgValArgSerSerSerGlyPr	1300
Db	3841	TGAGGGTGTGATGTCGTGTGACCCGATGCTCCAAAGAGAGACGGCAAGAACCTCAGGACC	3900
QY	1300	oProhIleGlnThhAlaAaPProLybAaGThhThraAaPProMetGlyGlyArgValaGlyAr	1320
Db	3901	CCCCCATGAGAGGGCGCGCCCCAAGCGCACTTAAGACATGATGAGGGCGCGGTGGCGAG	3960
QY	1320	gAlaIleSerSerAlaSerIleGlyGlyLeuMetGlyArgAlaIlePProProGluArgH	1340
Db	3961	AGCCATCTCTCCAGCCAGCATCGAAGGTCATGATGGCGCGGTGCATCCGCGGAGCGACGA	4020
QY	1340	sSerProhIlehIleuLybGlyGlnhIlehIleArgGlySerIleThhGlnGlyIlePr	1360
Db	4021	CAGCCCCACCATCTCAAGAGGACGACCAATCCGGGGTCCATTCACACAAAGGAAATCC	4080
QY	1360	aArgSerTyValGlnAlaGlnAlaPProTyLeuArgArgGlyAlaAlaLybLeuLybAr	1380
Db	4081	TCCGTCTCATGTGAAGGACACAGAGAGACTACGTGCGTGGGAGGCCAAAGCTCTTAAAGCG	4140
QY	1380	gGlyGlyThhPProPProPProPProPProSerArgAaPLeuThhGlyAlaTyLyThhGly	1400
Db	4141	GGAGGGGACGCTCCGCCCCACCGCCCTCAGCGGACCTGACCGAGGCTTACAAAGACGA	4200
QY	1400	nAlaLeuGlyPProLeuLybLeuLybProAlaHISGlyGlyLeuValaIleThhThraLybGly	1420
Db	4201	GGCCCTTGGGCCCTGTAAAGTGAAGCGCGCCATGAAGGGCGTGGTGGCCACGGTGAAGGA	4260
QY	1420	uAlaGlyArgSerIlehISGlyIlePProArgGlnGlyLeuArgHISThhPProGluLeuPr	1440
Db	4261	GGGGGGCGGCTCATTCATGATGATCCGGCGGAGAGACTCGGGACACAGCCCGGAGTGC	4320
QY	1440	oLeuAlaPProArgPProLeuLybGlyGlySerIleThhGlnGlyThhPProLeuLybTyAs	1460
Db	4321	CTGGGCCCCCGGGCGGTCAAGAGGGGCTCATTCACCAAGGACCCCGGCTCAAGTACGA	4380
QY	1460	pThhGlyAlaSerThhThhGlySerLybHISaAaPValaArgSerLeuIleGlySerPr	1480
Db	4381	CACCGGGCGGTCCACCATGCGTGCATAAAAGCAAGATGAGCTCCCTCATTCGCAAGCC	4440
QY	1480	oGlyArgThhPhePProPProValhISpProLeuAaPValaMetAlaAaPAlaArgAlaLeuGly	1500
Db	4441	CGGCGGAGACTTCCACACCGGTGACCGCGCTGATGTGAhTGGCGAGCGCCCGGAGCATGGA	4500

OY	1500	UARGAIAcYsTfYrgIuGIuSserLeuYsSerArYpProGIYThraIAaerSerGIYgl	1520
Db	4501	ACGgCCCTGCTACAGAGAGACCTGAAGAAGCCGGACCGGACCGCAAGCTCGGGGGG	4560
OY	1501	YSerIleAIArGIYAlAProValIleValProGIUleuGIYySPProArGIInSerPr	1540
Db	4561	CTTCATTGCGGGCGGGCCGCCGGCTATTGTGCTTGAGCTGGGTAAAGCCGGCGAGACC	4620
OY	1540	oleuThrTYrgIuAAPHISgIYAlAProPheAlAGIYHISleuDrProArGIYSerProVa	1560
Db	4621	CCTGACCTATGAGAGACCAGCGGGGACCTTTGGCGGCACTCCACGAGAGTTGGCCGT	4680
OY	1560	lThrMeArGIuProThrProArGIeInGIYSerIleuSerSerIYsAlase	1580
Db	4681	GACCAACGGGAGGCCACGACGGCGCGCTGCAGAGAGGACGCTTTCGTCCAGCAAGGATC	4740
OY	1580	rgIAAsPArGIYsLeuThrSerTrProArGIuIleAlaYsSerProHISerThVa	1600
Db	4741	CCAGGACCGAAGCTGAGCTGAGCGCTCGTGAATGCCAAGTCCCGCAGACACCT	4800
OY	1600	lProGIuHISISrProHISerProIleSerProTYrgIuHISleuLeuArGIYAlSerGI	1620
Db	4801	GCCGGAGACACCCACACCCCATCTGCGCTTAGAGACCTGCTCGGGCGTGAATGG	4860
OY	1620	YValAsPLeuTYrArGIserHISISleProLeuAlAPheAsPProHrYserIleProArGI	1640
Db	4861	CGTGACCTGTATCGCAGACCAATCCCTCGCTTCGACCCCACTCCATCAACCCCGCG	4920
OY	1640	YlIeProleuAsPAlaAlaAlaAlaTYrTYrLeuProArGIshleuAlAProAsPProTh	1660
Db	4921	CATCCCTTGAGACGACCGCTGCTACTACTGCCCCGACACCTGGACCCCAACCCAC	4980
OY	1660	rTYrProHISleuTYrProProTYrLeuIleArGIYrProAsPThraIAleuGI	1680
Db	4981	CTACCCGACCTGTACCAACCCCTACCTATCGCGGGCTAACCCGACAGGGCGGTGGA	5040
OY	1680	uAsnArGIInThrIleIleAsnAsPTYrIleThSerGIInGIeThSHISleAnThraI	1700
Db	5041	GAACCGGAGACCATCATCAATGACTACATCACCTCGAGAGATGACCAACCAACGGCGC	5100
OY	1700	aThraIAmerIAgInArGIaAsPMeLeuArGIYleuSerProArGIuSerSerIe	1720
Db	5101	CACCGCCATGGCCCGACGCTGATATCTGAAGGGCCCTCTGCGCCGCGAGTCTGCT	5160
OY	1720	uAlaIeuAsnTYrAlaAlaGIYProArGIYIleIeAsPleuSerGIInAlProHISle	1740
Db	5161	GGCACTCAACTACGTGCGGGGTCCCGAGGATATGCACTGTGCCAAGTGCACACCT	5220
OY	1740	uProValleuValProProThrProGIYThProAlaThraIAmeArGIeAlaIY	1760
Db	5221	GCCTGTGCTCGGCCCGGACACCGGACCCGACCCCACTGAGCCGCTGTGGCTTA	5280
OY	1760	rIeuProHrArIAProGIInProPheSerSerArGISerSerSerProIleuSerProGI	1780
Db	5281	CCTCCCAACCGGCCCAAGCCCTTATGAGCGGCAACGACCTCCACTCTCCAGG	5340
OY	1780	YGIYProThrHISleuThrIYsProThrThrThrSerSerSerGIuArGIuArGPAr	1800
Db	5341	AGGTCCACACACTTGAACAAACACACACACAGCTCTGTCGAGCGGAGCGAGACCG	5400
OY	1800	gAsPArGIuArGPArGPArGIuArGIuYsSerIleleuThrSerThrThrTh	1820
Db	5401	GGATCGAAGCGCGACCGGATCGGAGACGGGAAAGTCCATCTCCAGTCCACACGAC	5460
OY	1820	rValGIuHISAlAProIleTPArGPProGIYThrGIuInSerSerGIYSerSerGIYSe	1840
Db	5461	GGTGAGACACGACCATCTGAGACCTGTGTAACAAGAG-----ACACACGGCAG	5511
OY	1840	rSerGIYGIYGIYGIYSerSerSerArProAlaSerHISerHISAlaHISgInHIS	1860
Db	5512	CAGCGCGGGGGGTGGGGGAGAGACGCGGCCGCTCCCACTCCCATGCCACACGACA	5571
OY	1860	sSerProIleSerProArGIInGIAsPAlaIeInGIInArGPProSerValleuHISAs	1880

Db	5572	CTGCCCATCTCCCTCTGGACCAGATGCCCTTCAGAGAGACCAGTGTGTTCAACA	5631
QY	1880	nThgIyMeelysGIylelleThralaValGIuProSerLyseProThrValleuArgse	1900
Db	5632	CACAGGATGAAAGGGATCATCAACCGCTGTGGAGCCAGACACCACAGGTCTCTGAGATC	5691
QY	1900	rThSerThSerSerProValArgProAlaAlaThrPheProProAlaAlaThriAcysPr	1920
Db	5692	CACCTTCACCTCCCTCACCCCTGTGGCCCGAGCTGCACATTCACACCTGCACACCACTGACC	5751
QY	1920	oLeuGIyGIYThriLeuAspGIyValTYrProThrLeuMetGIuProValLeuLeuProLy	1940
Db	5752	ACTGGGGGGCACCTTCATGGGGGTCTACCTTACCTCATGGAGCCGTCTTGGTGCCTAA	5811
QY	1940	sGIuAlaProArgValAlaArgProGIuArgProArgAlaAspThrGIyHisAlaPheLe	1960
Db	5812	GGAGGGCCCCGGGTGCGCCGGCCAGAGCGAGCCCGGAGCAGACACGGGCATCCCTCTCT	5871
QY	1960	uAlaLyseProProAlaArgSerGIyLeuGIuProAlaSerSerProSerLyseGIySerGI	1980
Db	5872	CGCCAAGGCCCCACCCGCTCCGGGCTGGAGCCGCTCTCTCCCCAGCAAGGAGCTCGGA	5931
QY	1980	uProArgProLeuValProProValSerGIyHisAlaThrTlAlaArgThrProAlaLy	2000
Db	5932	GCCCCGGCCCCATGTGCTCTGTCTGTGGCCAGCCACCATGTGCCCGCACCCCTGGGA	5991
QY	2000	sAsnLeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSerAs	2020
Db	5992	GAACCTGGCACCTCACCCACCGCCAGCCCGGAGCCCGGCGGCACACTGTCCCTCGGCTCGGA	6051
QY	2020	pProHisArgGIuLyThrGIuSerLyseProPheSerTlLeGIuGIuLeuGIuLeuArgse	2040
Db	6052	CCCCCACCGGAAAAAGACTCAAACTAAACCTTTTTCATCCAGAGATCGGAACCTCGCTTC	6111
QY	2040	rLeuGIYThrThiGIySerSerTYrSerProGIuGIuValGIuProValSerProValSe	2060
Db	6112	TCTGGAGTTACACAGAGACAGACTACAGCCCGAAAGGGGTGAGCCGTCAAGCCCTGTGAG	6171
QY	2060	rSerProSerLeuThriHisApLySGIyLeuProLyHisLeuGIuGIuLeuAspLySe	2080
Db	6172	CTCACCCAGTGTGACCCACAGCAAGAGGGGCTCCCAACACACTGGMAAGAGCTCGACAAAGAG	6231
QY	2080	rHisLeuGIuGIyGIuLeuArgProLySGInProGIyProValLyseGIyGIyGIuAl	2100
Db	6232	CAACTGGAGGGGAGGCTGCGGCCAAGACAGCCGCGTGAAGCTTGTGGCGGGGAGGC	6291
QY	2100	AlaHisLeuProHisLeuArgProLeuProGIuSerGIuProSerSerProLeuLe	2120
Db	6292	CGCCCACTCCCAACACCGCGCGCTGCTCGAAGGACAGCCCTGTGCAGCCCGCTGCT	6351
QY	2120	uGIuThrAlaProGIyValLySGIyHisSGInArgValValThrLeuAlaGInHisTlSe	2140
Db	6352	CGAGACCGCCCAAGGGGTCAAAAGTCAACAGCGGGTGTCACTTGGCCACAGCACAATCAG	6411
QY	2140	rGIuValTlIleThrGInAspTYrThrArgHisHisPProGIuGIuLeuSerAlaProLeuPr	2160
Db	6412	TGAGGTATTCACACAGGACTACACCGGACACCAACACAGAGCTAGCCACCCCTGCC	6471
QY	2160	AlaAlaProLeuTYrSerPheProGIyAlaSerCysProValLeuAspLeuArgArgProPr	2180
Db	6472	CGCCCCCTCATCTCTCTCCCTGGGGGCGAGCTGCCCTGTGACCTCGCGCGCCAC	6531
QY	2180	oSerAspLeuTYrLeuProProProAspHisSGIyAlaProAlaArgLyseProHisSe	2200
Db	6532	CAGTACCTTCACTCCCGCCCGGACCAATGGGCCCCGGCCCTGTGCTCCCCCAG	6591
QY	2200	rGIuGIyGIyysArgSerProGIuProAsnLyThSerValLeuGIyGIyGIyAs	2220
Db	6592	CGAAGGGGGCAAGAGTCTCCAGAGCCAAACAAACAGCTGGTCTTGGGTGGTGTAGGA	6651
QY	2220	pGIyTlIleGIuProValSerProProGIuGIyMeThrGIuProGIyHisSerArgSerAl	2240

Db 6652 CGGATTTGAACCTGTGTCTCCACCGGAGGCGATGACGAGCCAGGCGCATCTCCGAGTGC 6711
 Qy 2240 aValTYrProleuLeuTYrAArgAPGlyGluGlnThrGluProSerAArgMetGlySerly 2260
 Db 6712 TGTGTACCCCTGCTGTATCCGGGATGGGGAACAGAGCCAGCCAGAGATGGGCTCCAA 6771
 Qy 2260 sSerProGlyAAsnThrSerGlnProProAlaPhePheSerlyLeuThrglySerAAsn 2280
 Db 6772 GTCTCCAGGCAACACGAGCCAGCCGCGCTTCTTCAAGCAAGCTGACCGAGCAATC 6831
 Qy 2280 fAlaMetVallySerlySerlyGlnGlnIleAsnlyLeuLeuAsnThrHisAsnArgAs 2300
 Db 6832 CGCCATGGTCAAGTCCAGAGAGAGATCAACAGAGAGTGAACACCCCAACCCGGA 6891
 Qy 2300 nGluProGlyUTyrAAsnIleSerGlnProGlyThrGlnIlePheAsnMetProAlaIleTh 2320
 Db 6892 TGAGCTGAAATACATATTCAGCCAGCTGGGACGGAAGTCTTCAATATGCCCCCAATAC 6951
 Qy 2320 rGlyThrGlyLeuMetThrTYrAArgSerGlnAlaValGlnGlnIleHisAsnThrAsn 2340
 Db 6952 CGGAACAGAGCTTATGACTATAGAGCAGCGGTGAGGAACATGCGACACCAACAT 7011
 Qy 2340 rGlyLeuGlnAlaIleIleArglyAlaLeuMetGlyLyserTYrAArgGlnTTPGluGlu 2360
 Db 7012 GGGGCTGAGGCGCATTAATTAGAAAGGCACTCATGGGTAAATATATACAGTGGGAAGATC 7071
 Qy 2360 rProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaI 2380
 Db 7072 CCGCCCTCAGCCGCAATGCTTTAAACCTCTAAAGCCAGTCCAGCCGCGCTGC 7131
 Qy 2380 aMetProIleThrAlaAlaAspGlyAArgSerAspHisThrLeuThrSerProGlyGly 2400
 Db 7132 TATGCCATAAACGCTGCTGACGAGCGAGAGACACACACTCACTCGCCAGGTGGCGG 7191
 Qy 2400 yGlyLyAlaAlaValSerGlyAArgProSerSerAArglyAlaLySerProAlaProG 2420
 Db 7192 CGGGAAGGCCAAGGTCTCTGCGACAGCCAGCAGCCGAAAGCCAAAGTCCCGGCGCGG 7251
 Qy 2420 yLeuAlaSerGlyAAspAArgProProSerAlaSerValHisSerGlnGlyAspGlyAs 2440
 Db 7252 CTTGGCATCTGGGAGCCGCGCACCTCTGCTCTCTAGTGACTCGAGGAGAGCTCA 7311
 Qy 2440 nArgArgThrProleuThrAsnArgValTTPGluAspAArgProSerSerAlaGlySerTh 2460
 Db 7312 CCGCGGAGCCGCTGTCACCAACCGCGTGGGAGAGACAGCGCTCTCCGAGGTTCCAC 7371
 Qy 2460 rProPheProTYrAsnProLeuIleMetAArgLeuGlnAlaGlyAlaMetAlaSerProP 2480
 Db 7372 GCCATTCCCTACAAACCCCTGATCATGCGGCTGACGCGGTGTCTATGCTCCAC 7431
 Qy 2480 oProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisAlaATTPAAspG 2500
 Db 7432 CCACCCGCGCTCCCGCGGAGCGGCGGCGCTGCGGCCCCACACACGCTGGAGCA 7491
 Qy 2500 uGluProLyProLeuLeuCySerGlnTYrGlnThrLeuSerAAspSerGln 2517
 Db 7492 GAGGCCCAACCACTGCTGCTGCGAGTACAGACACTCTCCAGACGCGAG 7543

RESULT 7

ACN39603
 ID ACN39603 standard; CDNA; 8533 BP.

ACN39603;

18-NOV-2004 (first entry)

Tumour-associated antigenic target (TAT) cDNA DNA325805, SEQ ID NO:3856.

Tumour-associated antigenic target; TAT; human; overexpression; cancer;
 colorectal cancer; lung cancer; ovarian cancer; liver cancer;
 central nervous system cancer; bladder cancer; pancreatic cancer;
 cervical cancer; melanoma; leukaemia; hybridisation probe;

KM chromosome identification; chromosome mapping; gene mapping;
 KM gene therapy; cytostatic; gene; ss.
 XX Homo sapiens.
 XX MO2004030615-A2.
 XX 15-APR-2004.
 XX 29-SEP-2003; 2003MO-US028547.
 XX 02-OCT-2002; 2002JUS-0414971P.
 XX (GERTH) GENENTECH INC.
 XX Mu TD, Zhang Z, Zhou Y;
 XX WPI; 2004-347921/32.
 XX New tumor-associated antigenic target polypeptides and nucleic acids,
 PT useful in preparing a medicament for treating or detecting a
 PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
 PT prostate cancer or tumor.
 PS Claim 1; SEQ ID NO 3856; 7273bp; English.

CC The invention relates to human tumour-associated antigenic target (TAT)
 CC polypeptides, and their related nucleic acids. The TAT polypeptides are
 CC overexpressed in cancer tissues compared to normal tissues, and may thus
 CC serve as effective targets for the diagnosis and treatment of cancer in
 CC mammals. The invention also relates to nucleic acid and polypeptide
 CC sequences at least 80% identical to the TAT nucleic acids and
 CC polypeptides; expression vectors and host cells comprising a TAT nucleic
 CC acid, an antibody specific for a TAT polypeptide; a peptide or organic
 CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
 CC TAT polypeptide; and methods and compositions for the treatment or
 CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
 CC antibodies, antagonists, binding molecules and compositions are useful
 CC for diagnosing or treating a cell proliferative disorder associated with
 CC increased TAT expression, particularly cancers such as breast cancer,
 CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
 CC cancer, pancreatic cancer, cervical cancer, cancers of the central
 CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be
 CC used as hybridisation probes, in chromosome and gene mapping, in
 CC chromosome identification and in gene therapy. The present sequence
 CC represents a TAT nucleic acid of the invention

SQ Sequence 8533 BP; 1836 A; 3034 C; 2528 G; 1135 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0 Length: 8533
 Score: 13119.50 Matches: 2506
 Percent Similarity: 99.56% Conservative: 1
 Best Local Similarity: 99.52% Mismatches: 7
 Query Match: 99.28% Indels: 5
 DB: 13 Gaps: 1

US-09-522-753-5 (1-2517) X ACN39603 (1-8533)

Qy 1 MetSerGlySerThrGlnLeuValAlaGlnThrTYrAArgAlaThrGluProAArgTYrPro 20
 Db 2 ATGTGGGATCCACACAGCTGTGGCAGACAGCTGGAGGGCCACTGAGCCCCCTACCGG 61
 Qy 21 ProHisSerLeuSerTYrProValGlnIleAlaArgThrHisThrAAspValAlaGlyLeu 40
 Db 62 CCCACAGCTTCTTCTAACCAGTGCAGATGCGCGGACGACACGAGCGTGGGCTCTCG 121
 Qy 41 GluTYrGlnHisHisSerAArgTYrAlaSerHisSerSerProGlySerIleIleGln 60
 Db 122 GATTACAGACCACTCCCGGACTATGCTCCCACTGTGCGCGGCTCATCATCAG 181
 Qy 61 ProGlnAArgAArgProSerLeuLeuSerGlnPheGlnProGlyAAsnGluAArgSerGln 80

Db 182 CCCACGGCGAGGCCCTCCCTGCTGTGAGTTCAGATCCCGGAATGACGCTCCAG 241
Qy 81 GluLeuHisIleuAspProGluSerHisSerTyrIleuProGluLeuGlyIleuSerGluMet 100
Db 242 GAGCTCACCTCGGCGACAGAGTCCCATCTACTGCGGAGCTGGGAAAGTCAGAGATG 301
Qy 101 GluPheIleGluSerIleAspProAlaGluLeuGluLeuLeuProAspProIleuLeuAspPro 120
Db 302 GAGTTATTGAAAGCAAGCGCCCTCGGCTGAGAGCTGCTGCTGACCCCTGCTGCCACCG 361
Qy 121 SerProIleuLeuAlaThrGlyGlnProAlaGlySerGluAspIleuThrIleAspAspSer 140
Db 362 TCACCCCTGTGGCCACGGGCGAGCCCTGCGGAGTCTGAAGACCTCAACAAAGACCGTAG 421
Qy 141 LeuThrGlyIleuSerGluProValSerProProSerProProHisThrAspProGluLeu 160
Db 422 CTGACGGGAGAGTGAACCGGTGTCTCCCCACAGCCCCCGGACACTGACCTTGAGCTG 481
Qy 161 GluLeuValProProAlaGluSerIleuSerGlyGluLeuIleGluAsnMetAspArgValAsp 180
Db 482 GAGCTGTGCGCCGACAGCTGTCCAAAGAGAGCTGATCCAGAACATGGACCCGCTGGAC 541
Qy 181 ArgGluIleThrMetValGluGlnGlnIleIleTyrAspGluAsnArgIleuValGluVal 200
Db 542 CGAGAGATGACCATGGTGAAGACAGAGATCTTAAAGCTGAAGAAAGACAGACAGCTG 601
Qy 201 GluGluGluValAlaAlaIleAspProProGluProGluIleuProValSerProProIleGlu 220
Db 602 GAGGAGAGAGGCTGCCAAGCGCGCCGAGCCCTGAGAAAGCCCTGTACCGCGCCCATCGAG 661
Qy 221 SerIleHisIleArgSerIleuValGlnIleIleTyrAspGluAsnArgIleuValGluVal 240
Db 662 TCGAAGCACCGCAGCTGTGTGAGATCATCTAAGAGAGAACCGGAAGAGCTGAAGCT 721
Qy 241 AlaHisArgIleIleuGluGluIleuGlyProGluIleValGluLeuProIleuTyrAsnGlnPro 260
Db 722 GACATCGAGATTCGGAAGGCTGGGAGCCCGGAGTGAAGCTGCGCTGTAACAACGAGCC 781
Qy 261 SerAspThrArgGlnTyrHisIleGluAsnIleIleValAsnGlnAlaMetArgIleuValLeu 280
Db 782 TCCGACACCCGCGAGTATCTAGGAACATCTAAATTAACCAAGCGCATCGCGAAGAAAGCTA 841
Qy 281 IleLeuTyrPheIleAspArgIleuAsnHisAlaArgIleGlnIleTyrGlnIleuPheCysGln 300
Db 842 ATCTTGTACTTCAAGAGAGGAATCAAGCTCGGAAACAATGGAGACGAAGATTCTGCCAG 901
Qy 301 ArgTyrAspGlnIleuMetGluAlaLeuGluIleuValGluArgIleGluAsnAsnPro 320
Db 902 CGCTATGACCAAGCTCATGGAGGCTGGGAGAAAGAGTGAAGCGCATCGGAACAACCCC 961
Qy 321 ArgArgArgAlaIleuSerIleuValArgGluTyrTyrGluIleuSerGlnPheProGluIle 340
Db 962 CGCGCGCGGCGCCAGAGAGACAGAGTGTCCGAGTACTACAGAAAGCAAGTTCCCTGAGATC 1021
Qy 341 ArgIleuSerGlnArgIleuGlnIleuArgMetGlnSerArgValGlyIleuArgIleuSerGly 360
Db 1022 CGCAAGCAGCGGAGCTGACAGAGCGCATGCAAGACAGGAGTGGCCACGCGGCGCAATGG 1081
Qy 361 IleuSerMetSerAlaAlaIleuSerGluHisIleuValSerGluIleIleAspGlyLeuSer 380
Db 1082 CTGTTCATGTTCGGCCCGCCGACGAGCAAGAGTGTCAAGAGATCTGATGCTCTCTCA 1141
Qy 381 GluGlnGluAsnIleuGlnIleuSerGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400
Db 1142 GAGCAGAGAACTTGAAGAAAGCATGCGCCAGCTGGCCGTGATCCCGCCCATGCTGTAC 1201
Qy 401 AspAlaAspGlnIleuArgIleuIleuPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420
Db 1202 GACGCTGACAGAGGAGCATCAAGTTCATCAATGACAGCGCTTATGGCGGACCCCATG 1261
Qy 421 LysValTyrIleAspArgGlnValMetAsnMetTyrSerGluGlnIleuSerGluThrPhe 440
Db 1262 AAGGTGTACAAAGACCGCAGTCAATGAACATGTGAGTGAAGAGAGAGAGACCTTC 1321

Qy 441 ArgGluIlePheMetGlnHisProIleAsnPheGlyLeuIleAlaSerPheIleuGluArg 460
Db 1322 CGGAGAAAGTTCATGACAGATCCCAAGAACTTGTGCGCTGATGCATCATTCCTGGAGAG 1381
Qy 461 LysThrValAlaGluCysValIleuTyrTyrTyrIleuThrIleuIleuAsnGlyuAsnTyrLys 480
Db 1382 AAGACATGGCTGAGAGCTCTCTCTATTAATCACTGACTTAAGAAAGATAGAATCTATTAAG 1441
Qy 481 SerIleuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGln 500
Db 1442 AGCTGTGTAAGACGAGATCTATCGGCGCCCGGCAAGAGCAGCAGACGACACAGCAG 1501
Qy 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
Db 1502 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1561
Qy 521 AspGluIleuGluIleuSerGluValAlaGluIleuGluIleuProGluIleuValGluAsn 540
Db 1562 GATGAGAAAGAGAGAAAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1621
Qy 541 AspIleuGluAspIleuLeuIleuSerGluIleuThrAspAspThrSerGlyGluAspAspGlu 560
Db 1622 GACACAGAAAGACTCTCTCAAGAGAGAGAGAGAGACGACCTCAAGGAGAGAGAGAGAG 1681
Qy 561 LysGluAlaValAlaSerIleuSerGlyArgIleuThrAlaAsnSerGlnIleuArgArgIle 580
Db 1682 AAGAGAGCTGTGGCTCCAAAGCGCGCAAACTGCCAACAGCAGCGAGAGAGAGAGAG 1741
Qy 581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600
Db 1742 CGCATCAACCGCTCATAGGCTAATGAGGCCAACAGAGAGAGAGAGAGAGAGAGAGAG 1801
Qy 601 SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTyrThrGluGluGluMet 620
Db 1802 AGCGCCAGCTGGCTCTCAAGAGCTGAATGAAGTTCCTGCGACAGAGAGAGAGAGAG 1861
Qy 621 GluThrAlaIleuIleuSerGluLeuGlnIleuIleuArgAsnTyrSerAlaIleuArgMet 640
Db 1862 GAAACAGCAAGAAAGTCTCTCGGAACAGCGCGCAATGGTGGCGCATCGCGCGGATG 1921
Qy 641 ValGlySerIleuThrValSerGlnCysIleuAsnPheTyrPheAsnTyrIleuValArgGln 660
Db 1922 GTGGGCTCCAAAGCTGTGCGAGTGTAAAGACTTCTTCAACATCAAGAAAGAGAGCAG 1981
Qy 661 AsnIleuAspGluIleuGlnGlnHisIleuSerIleuMetGluIleuSerGluArgAsnAlaArg 680
Db 1982 AACCTGATGAGATCTTGACAGACAGCAAGCTGAAAGATGAAGAGAGAGAGAGAGAG 2041
Qy 681 ArgIleuIleuValSerAlaProAlaAlaIleuSerGluGluAlaIlePheProProVal 700
Db 2042 AGGAGAGAAAGAAAGCGCGCGCGCGCGCGCAGAGAGAGAGTGCATTCGCGCGCGGTG 2101
Qy 701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluIleuMetValGluGlu 720
Db 2102 GAGGATGAGAGATGAGAGCGTTCGGGCTGAGACCGGAAAGAGAGAGAGTGTGAGAGAG 2161
Qy 721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAla 740
Db 2162 GCTGAAGCTTCAATGCTCTGAGAAATGAGTGTCCAGAGGAGAAAGCATGTGCGCCAGGC 2221
Qy 741 ThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLys 760
Db 2222 ACTGTCAACAACGCTCAACACACAGAGACATCCCTCTCTCAACATGAGCGCCCAAG 2281
Qy 761 AspThrGlyGlnAsnGlyProIleuProProAlaThrIleuGlyAlaAspGlyProProPro 780
Db 2282 GACACAGGCGCAAAATGGCCCAAGCCCGCAGCACTGTGGGCGCGGACCGGCGCACCCCA 2341
Qy 781 GlyProProThrProProArgArgThrSerArgAlaProIleGluProThrProAlaSer 800
Db 2342 GGGCCACCCACCCACACCGAGAGACATCGCGGC -CCCACTGAGCCCAACCCCGGCTTC 2400

QY 800 rgIuaIaThrGlyAaProThrProProProAlaProProSerProSerAlaProProPr 820
| | | | |
Db 2401 TGAAGCCACCGAGCCCTTACGCCCCACGACCCCATCGCTTGCACCTCTCC 2460
QY 820 ovaIvaIProLybGluGluLybGluGluGluThraIaAlaIaProProValGluGluG 840
| | | | |
Db 2461 TGGGTCTCCCAAG 2520
QY 840 ygluGluGluLybProProAlaIaGluGluGluLeuAlaValAspThrGlyLybAlaGluG 860
| | | | |
Db 2521 GGAGAGACAAAGCCCTCCGCGCTTGAGAGAGCTGGAGTGGACACAGAGAAAGCCGAGGA 2580
QY 860 uproValLybSerGluCybThrGluGluGluGluGluGluGluGluGluGluGluGluG 880
| | | | |
Db 2581 GCCCGTCAAG 2640
QY 880 agIuaIaAlaGluAlaThraIaGluGluGluAlaLeuLybAlaGluLybGluGluGluG 900
| | | | |
Db 2641 GGAAGCCGCTGAGGCGCCAGCGCGGCGCTCAAGGAGAGAGAGAGAGAGAGAGAGAG 2700
QY 900 rgIyArGAlaThThraIaLybSerSerGlyAlaProGluAspSerAspSerAlaTh 920
| | | | |
Db 2701 CGGAGAGGACCAACAG 2760
QY 920 rCybSerAlaAspGluValaAspGluAlaGluGluGluGluGluGluGluGluGluGlu 940
| | | | |
Db 2761 CTGACGTGACAG 2820
QY 940 oArGProSerLeuLeuThraProThrGlyAaProAlaAlaAlaSerProGluLyPr 960
| | | | |
Db 2821 AAGGCCAGGCTCTCAACCCGAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2880
QY 960 oLeuAspLeuLybGluLeuLybGluGluGluGluGluGluGluGluGluGluGluGlu 980
| | | | |
Db 2881 ACTGAGCTTAAG 2940
QY 980 svaIhIaGluProProAlaGluAlaAlaProThraIaProThraIaProAlaProPr 1000
| | | | |
Db 2941 AGTCCATGAGCCCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3000
QY 1000 oProProGluAlaAlaGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 1020
| | | | |
Db 3001 GGCACCGCAAAAGCTGAG 3060
QY 1020 gglLybSerArGSerProAlaProAlaAlaAspLybGluAlaPheAlaIaGluAlaG 1040
| | | | |
Db 3061 GGGCAAG 3120
QY 1040 nLybLeuProGlyAaProProCybThraIaSerGlyLeuProPheProValProProAr 1060
| | | | |
Db 3121 GAAGCTGCTGGGAG 3180
QY 1060 ggluValIleLybAlaSerProhIaIaProAspProSerAlaPheSerTyraIaProPr 1080
| | | | |
Db 3181 TGAAGGATCAAG 3240
QY 1080 ogIyHIAProLeuProLeuGluLybLeuIaAspThraIaArGProValLeuProArProPr 1100
| | | | |
Db 3241 TGGTCAACCACTGCTCCCTGAGCTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3300
QY 1100 oThraIleSerAspProProProLeuIleSerSerAlaLybAlaPheProSerValLeuGluAr 1120
| | | | |
Db 3301 CACCACTCAACAG 3360
QY 1120 ggluIleGlyAlaIleSerGluGlyMetSerValGluLeuHleValProTySerGluH 1140
| | | | |
Db 3361 GCAAAAG 3420
QY 1140 sAlaLybAlaProValGlyProValThraMetGlyLeuProLeuProMetAspProLybLy 1160
| | | | |
Db 3421 TGGCAAG 3480
QY 1160 sLeuAlaProPheSerGlyValLybGluGluGluGluGluGluGluGluGluGluGluGlu 1180

Db 3481 GCTGAGACCTTACGAG 3540
QY 1180 oProGluSerLeuGlyAlaProThraIaGluGluGluGluGluGluGluGluGluGluGlu 1200
| | | | |
Db 3541 ACCGAG 3600
QY 1200 uGlySerValProGlyValLybSerIleThraLybGlyIleProSerThraArGValProSerAr 1220
| | | | |
Db 3601 GGGCTCACTTCCGAG 3660
QY 1220 pSerAlaIleThraIaGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 1240
| | | | |
Db 3661 CAGGAGCATACATACCGGAG 3720
QY 1240 sglYThraIleThraIaGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 1260
| | | | |
Db 3721 GGGCAGCATACAGAGATCAATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3780
QY 1260 uAspSerLeuProLybGlyHleIaValIleTyrgluGlyLybAlaIaValLeuSerTy 1280
| | | | |
Db 3781 GAGAGAGCTCCAG 3840
QY 1280 rgIuGlyGlyMetSerValThraGlySerLybGluAspGlyArGSerSerSerGlyPr 1300
| | | | |
Db 3841 TGAAGGTGGATGTCTGTGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3900
QY 1300 oProHleGluThraIaAlaProLybArGThraIaPheMetMetGluGlyArGlyAr 1320
| | | | |
Db 3901 CCCCAGATGAG 3960
QY 1320 gAlaIleSerSerAlaSerIleGluGluGluGluGluGluGluGluGluGluGluGluGlu 1340
| | | | |
Db 3961 AGCATCTCTTCAAG 4020
QY 1340 sSerProHleHleLybGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 1360
| | | | |
Db 4021 CAGCCCAACCACTCAAG 4080
QY 1360 oArGSerTyraIaGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 1380
| | | | |
Db 4081 TCGGTCTTACGTGAG 4140
QY 1380 ggluGlyThraProProProProProProProProProProProProProProProPro 1400
| | | | |
Db 4141 GAG 4200
QY 1400 nAlaLeuGlyProLeuLybLeuLybProAlaHleGluGluGluGluGluGluGluGluGlu 1420
| | | | |
Db 4201 GGCCTGGGAG 4260
QY 1420 uAlaGlyArGSerIleHleGluIleProArGluGluGluGluGluGluGluGluGluGlu 1440
| | | | |
Db 4261 GGGGAG 4320
QY 1440 oLeuAlaProArGProLeuLybGluGluGluGluGluGluGluGluGluGluGluGluGlu 1460
| | | | |
Db 4321 CTTGGCCGAG 4380
QY 1460 pThraIaIaSerThraIaGlySerLybLybAlaPheValaIaSerSerIleGlySerPr 1480
| | | | |
Db 4381 CACGAG 4440
QY 1480 ogIyArGThraPheProProValHleProLeuAspValMetAlaAspAlaAlaGluGlu 1500
| | | | |
Db 4441 CGGCGAG 4500
QY 1500 uArGAlaCybTyrgluGlySerLeuLybSerArGProGlyThraIaSerSerSerGlyG 1520
| | | | |
Db 4501 AGCTGCTCTAG 4560
QY 1520 ySerIleAlaArGlyAlaProValIleValProGluLeuGlyLybProArGluGluSerPr 1540
| | | | |

Dh 4561 CTCATTGGCGGCGGCCCGGTCATTGTGCTGAGCTGGGTAAGCCGCGAGAGCCC 4620
Qy 1540 oLeuThrTyGluAspHisGlyValaProPheAlaGlyHisLeuProArgGlySerProVa 1560
Db 4621 CCGAGCTTAAGAGCAACGCGGACCCCTTGCCGGGCACTCCAGAGGTTGCGCCGT 4680
Qy 1560 LThrMetArgGluProThrProArgLeuGlnGlySerLeuSerSerSerLyAlaSe 1580
Db 4681 GACCAAGCGGAGAGCCCAAGCGCGCTGCGAGAGGAGCGCTTTCGTCAGCAAGGCATC 4740
Qy 1580 rGlnAspArgLyLeuThrSerThrProArgGluAlaAlaValSerProHisSerThrVa 1600
Db 4741 CAGAGACGGAACCTGAAGTGAAGCGCTGTAAGATGCGCAAGTCCCGAGACGACCGT 4800
Qy 1600 LProGlnHisProHisProHisProHisProHisProHisProHisProHisPro 1620
Db 4801 GCCCGAGCAACCAACACCCCATCTGCGCTTAAGACCTGTCTTGCGGGCGCTGAGTGG 4860
Qy 1620 yValaAspLeuTyArgSerHisLeuProLeuAlaPheAspProThrSerLeuProArgG 1640
Db 4861 CGTGGACCTGTATCGGAGCGACATCCCGCTGCGCTTCGACCCCATCCATACCCGCGG 4920
Qy 1640 yLLeuProLeuAspAlaAlaAlaAlaTyTyTyLeuProArgHisLeuAlaProAspPro 1660
Db 4921 CATCCCTCTGAGAGCAAGCGCTGCTTAAGTCTGCGCCGAGACCTGAGCCCAACCCAC 4980
Qy 1660 rTyProHisLeuLeuTyProProTyTyLeuLeuArgGlyTyTyProAspThrAlaLeuG 1680
Db 4981 CTACCGGCACTGTACCACTTACCTTACCTGCGGGCTACCCGAGCAAGCGCGGCTGGA 5040
Qy 1680 uAsnAspGlnThrIleIleAsnAspTyTyleThrSerGlnGlnMetHisHisAsnThrAl 1700
Db 5041 GAACCGGAGACCATCATCAATCAATCACTCGAGCAGATGACACCAACCGCGC 5100
Qy 1700 aThrAlaMetAlaGlnAlaArgAlaAspMetLeuArgGlyLeuSerProArgGlnSerSe 1720
Db 5101 CACCGCATGAGCCGAGAGCTGATATGCTGAGAGGCTCTGCGCCGAGAGCTCTGGCT 5160
Qy 1720 uAlaLeuAsnTyAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLe 1740
Db 5161 GGGACTCAACTAGCTGCGGGTCCCGAGGCACTGAGACTGTCTCCAGATGCAACACT 5220
Qy 1740 uProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTy 1760
Db 5221 GCGTGTGCTGTGCTCCCGGACACAGGACCCGAGCACCAGCATGAGACCGCTTCCCTA 5280
Qy 1760 rLeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerPro 1780
Db 5281 CCTCCCAACCGGCGCCAGGCTTCAGACAGCGCCCAAGCAAGCTCCCACTTCCCAAGG 5340
Qy 1780 yGlyProThrHisLeuThrTyAspProThrThrThrSerSerSerGlyuArgGlyuArgAspAr 1800
Db 5341 AGGTCCAAACACTTGAACAAACCAACCAACCAAGTCTCTGTCGAGCGGAGGAGAACCG 5400
Qy 1800 gAspArgGlyuArgAspArgAspArgGlyuArgGlyuArgGlyuArgGlyuArgGlyuArg 1820
Db 5401 GGATTCAGAGCGGAGACCGGATCGGAGCGGAGAAAGTCCATCTCACTCCACGACGAGC 5460
Qy 1820 rValGlnHisAlaProIleThrArgProGlyThrArgGlnSerSerSerGlySerSerGlySe 1840
Db 5461 GGTGGAGCAAGCACTCTGAGACCTGTATACAGAGCAG-----AGCAGCGGAG 5511
Qy 1840 rSerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHis 1860
Db 5512 CAGCGCGGGGGGGGGGAGAGAGCGCGCGCTCCCACTCCCAATGCCACAGCA 5571
Qy 1860 sSerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAla 1880
Db 5572 CTCGCCCATCTCCCTCGGACCCAGAGATGCCCTCCAGCAGAGAACCAAGTGTCTTCAAA 5631
Qy 1880 nThrGlyMetLyGlyIleIleThrAlaValGlnProSerSerLyProThrValLeuArgSe 1900
Db 5632 CACAGCATGAAGGATATCATCAACGCTGTGAGCCCAAGCAAGCCCAAGTCTCTGAGGTC 5691

Qy 1900 rThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisGlySer 1920
Db 5692 CACCTCACTCTCCACCGGTTGCGCGGCTGCGACATTCACCTGACCAACCACTGCCC 5751
Qy 1920 oLeuGlyGlyThrLeuAspGlyValTyProThrIleMetMetGluProValLeuLeuProLy 1940
Db 5752 ACTGGGCGGACCCCTGATGGAGGTCTTACCTTACCTCAATGAGCCCGCTGTGTGCCAA 5811
Qy 1940 sGluAlaProArgValAlaArgProGluArgProAlaArgAlaAspThrGlyHisAlaPheLe 1960
Db 5812 GAGAGCCCCCGGGGTGCGCGGCGAGAGCGGCCGAGACAGACAGCGGCAATCCCTTCT 5871
Qy 1960 uAlaTyProProAlaArgSerGlyLeuGluProAlaSerSerProSerLySerGlySerG 1980
Db 5872 CGCAAGCCCCCAGCGCGCTCGGGCTGAGGCGCGCTCTCCCGCAGCAAGGCTCGGA 5931
Qy 1980 uProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaTy 2000
Db 5932 GCCCGGCGCCCTAGTGCTCTGTCTGTGCAAGCGCACCATGCGCGGACCCCTTCGAA 5991
Qy 2000 sAsnLeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSerAs 2020
Db 5992 GAACCTCGACCTCAGCAAGCGGACCGGAGCCCGCGGCACTGCTCGGCTCGGA 6051
Qy 2020 pProHisArgGlyuArgThrGlnSerLyAspProPheSerIleGlnGluLeuGlnLeuArgSe 2040
Db 6052 CCGGCAACCGGAGAAAGACTCAAAAGTAAACCTTTTCATCCAGGAATGGAATCCGTTTC 6111
Qy 2040 rLeuGlyTyHisGlySerSerTySerProGlnGlyValGlnProValSerProValSe 2060
Db 6112 TCTGGATTACACGCGAGAGCTACAGCCCGCAAGGGGTGAGCCGTGAGCTGTGAG 6171
Qy 2060 rSerProSerLeuThrHisAspLySerGlyLeuProLyHisLeuGlnGluLeuAspLySe 2080
Db 6172 CTCACCACTCTGACCCCAAGCAAGGGGCTCCCAAGCACTGGAAGACTGCAAGAG 6231
Qy 2080 rHisLeuGlnGlyGluLeuArgProLySerGlnProGlyProValLyLeuGlyGlyGlyuAl 2100
Db 6232 CCACTGAGAGGGAGCTCGGCGCCCAAGCAGCAGGCGCCGTGAAGCTTGCGCGGAGGCG 6291
Qy 2100 sAlaHisLeuProHisLeuArgProLeuProGlnSerGlnProSerSerSerProLeuLe 2120
Db 6292 CGCCCACTCCCAACATGCGGCGCTGTGCTGAAGCCAGCTGTCCAGCGCGGTGCT 6351
Qy 2120 uGlnThrAlaProGlyValLyGlyHisGlnArgValValaThrLeuAlaGlnHisIleSe 2140
Db 6352 CAGAGCGGCCCAAGGGGTCAAAAGTCAACAGCGGGGTGTCACCTTGCCAGCAACTCA 6411
Qy 2140 rGluValIleThrGlnAspTyThrArgHisHisProGlnGlnLeuSerAlaProLeuPr 2160
Db 6412 TGAAGTCATCACAGAGTACACACCGGACCAACCAAGAGCTGAGCGACCCCTGCC 6471
Qy 2160 oAlaProLeuTyTySerPheProGlyValaSerTySerProValLeuAspLeuArgProPr 2180
Db 6472 CGCCCCCTCTACTCTTCTCTGAGGCGAGCTGCCCGTCTGAGACTTCCGCGCCAC 6531
Qy 2180 oSerAspLeuTyTyLeuProProProAspHisGlyAlaProAlaArgGlySerProHisSe 2200
Db 6532 CAGTACCTTACCTTCCGCGCCCGGACCAATGTGCGCGCGGTCTCCCGCCACAG 6591
Qy 2200 rGlnGlyGlyLyArgSerProGluProAsnLyThrSerValLeuGlyGlyGlyuAs 2220
Db 6592 CGAAGGGGAGAGGCTCCAGAGCAAGCAAGAGTGGCTTGGGTGGTGTAGGA 6651
Qy 2220 pGlyIleGluProValSerProProGlnGlyMetThrGlnProGlyHisSerArgSerAl 2240
Db 6652 CGGATTTGAACCTGTGTCCCAACCGAGGGCAATGACGAGGCAAGGCACTCCGAGGTGC 6711
Qy 2240 aValTyProLeuLeuTyArgAspGlyGlyGlnThrGlnProSerArgMetGlySerLy 2260
Db 6712 TGTGTACCGGCTGTACTACCGGAGATGGGAAACAGACGAGCCCAAGAGATGGGCTCCA 6771

QY 2260 sSerProGlyAAsnThrSerGlnProProAlaPhePheSerIybluThrgIuSerAAsnSe 2280
DB 6772 GTCTCCAGGCAACACCGCCAGCGCTTCTTGAGCAAGCTGACCGAGCAACTC 6831
QY 2280 rAlaMetValIySerIybluGlnGluIleAsnIybluLeuAAsnThrIleAAsnAAGAB 2300
DB 6832 CGCCATGTCAGAGTCAGAGAGCAAGAGATCAACAGAGAGTCGAAACCCCAACCCGAA 6891
QY 2300 nGluProGluIyTranIleSerGlnProGluIyThGluIlePheAAsnMetProAlaIleTh 2320
DB 6892 TGAAGCTGATACATATTCAGCCAGCTGGAGCGAGATCTTCAATATGCCCCGCTACAC 6951
QY 2320 rGlyIyThGlyLeuMetThrIyTrArgSerGlnAlaValGlnGluIleIleAAsnThrAAsnMe 2340
DB 6952 CGGAAACAGGCTTATGACTATAGACCGAGCGCTGAGGAAATGTCAGACACCAACAT 7011
QY 2340 tGlyIyLeuGluAlaIleIleIybluGlyAlaLeuMetGlyIyIyTrArgGlnIyTrGluIySe 2360
DB 7012 GGGGCTGGAGGCCATTAATTAGAAAGGCACTCATGGCTAAATATGACAGTGGAGAGATC 7071
QY 2360 rProProLeuSerAlaAAsnAlaPheAAsnProLeuAAsnAlaSerAlaSerIeuProAlaAl 2380
DB 7072 CCGCGCCCTAGCGCCCAATGCTTTAACTCTTGAATGCCAGTCCAGCTGCCCGCTGC 7131
QY 2380 aMetProIleThrAlaAlaAAspGlyArgSerAAspIleThrLeuThSerProGlyIyG 2400
DB 7132 TATGCCCATTAACCGCTGCTGAGCGAGGATGACCACTCACTCGCGCAGGTGGCGG 7191
QY 2400 yGlyIyAlaIyAlaIyValSerGlyIyArgProSerSerArgIyAlaIySerProAlaProG 2420
DB 7192 CGGAAAGCCAGAGTCTCTGGCAGAACCCAGACGCCGAAAGCCGAAAGTCCCGCGCCGCGG 7251
QY 2420 yLeuAlaSerGlyAAspArgProSerValSerValHisSerGluGlyIyAAspCysAs 2440
DB 7252 CTGGGACTCTGGGAGACCGCCCACTTGTCTCTCTGCTGACTCGAGGGAAGACTCAA 7311
QY 2440 nArgIyThrProLeuThrAAsnArgValIyTrGluAAspArgProSerSerAlaGlySerTh 2460
DB 7312 CCGCGCGAGCGCGCTCAACACCGCGTGGAGAGACAGGCCCTCGTCCGAGGTTCCAC 7371
QY 2460 rProPheProIyTrAAsnProLeuIleMetArgIyLeuGlnAlaGlyValMetAlaSerPro 2480
DB 7372 GCCATTCCCTACCAACCCCTGATCATGCGGCTGACGCGGAGTGTCACTGCTCCACC 7431
QY 2480 oProProGluIyLeuProAlaGlySerGlyProIyAlaGlyProHisIleIleIyTrAAsp 2500
DB 7432 CCCACCGGCGCTCCCGCGGAGCGGCGGCTGCTGCGCTGCGCCACCAACCGCTGGAGCA 7491
QY 2500 uGluProIyAAspProLeuLeuCysSerGlnIyTrGluThrLeuSerAAspSerGlu 2517
DB 7492 GAGGCCCAAGCACTGCTGCTGCTGCAAGTACAGAACACTCTCCGACAGCGAG 7543
RESULT 8
ACa62250
ID ACa62250 standard; cDNA; 7521 BP.
XX
AC ACa62250;
XX
DT 12-AUG-2003 (first entry)
XX
DE Human nuclear receptor corepressor SMRte coding region cDNA.
XX
KM Human; ss: gene; SMRte: nuclear receptor corepressor; gene therapy;
XX
OS tissue typing; cancer.
XX
OS Homo sapiens.
XX
XX
FH Key 1. location/Qualifiers
FT CDS 1. .7524
FT FT /tag= a
FT FT /partial
FT FT /product= "SMRte"
FT FT /note= "No stop codon given"

XX US2003027137-A1.
PN 06-FEB-2003.
XX
PD 27-MAR-2001; 2001US-00819104.
XX
PF 29-MAR-2000; 2000US-0193138P.
XX
PR (CHEN/) CHEN J D.
XX
PA
XX
PI Chen JD;
XX
DR WPI: 2003-466139/44.
XX
DR P-PDB; ABU61812.
XX
PT New SMRte proteins and nucleic acids, useful in gene therapy, predictive
PT medicine, therapeutic or prophylactic treatment, chromosome mapping,
PT tissue typing and in forensic biology.
XX
PS Claim 2; Page 48-56; 90pp; English.
XX
CC The invention relates to an isolated SMRte nucleic acid molecule. The
CC nucleic acids are useful in gene therapy, as hybridisation probes for
CC identifying SMRte-encoding nucleic acid molecules and as primers for
CC amplifying of SMRte nucleic acid molecules. The polypeptides are useful
CC as immunogens to raise anti-SMRte antibodies. The SMRte molecules are
CC useful as targets for discovering and developing modulating agents to
CC regulate a variety of cellular processes, in screening assays, in
CC predictive medicine, in therapeutic or prophylactic treatment, in
CC chromosome mapping, tissue typing and in forensic identification of a
CC biological sample. Modulators of SMRte are useful for treating or
CC preventing a condition associated with aberrant SMRte protein or nucleic
CC acid expression or activity, such as cancer. The present sequence
CC represents the human nuclear receptor corepressor SMRte coding region
CC cDNA
XX
SQ Sequence 7521 BP; 1635 A; 2728 C; 2212 G; 946 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 0 Length: 7521
Score: 12978.00 Matches: 2487
Percent Similarity: 98.57% Conservative: 3
Best Local Similarity: 98.46% Mismatches: 9
Query Match: 98.21% Indels: 28
DB: 8 Gaps: 3
US-09-522-753-5 (1-2517) x ACa62250 (1-7521)
QY 1 MetSerGlySerThrgIuLeuValAlaGlnThrTrpArgAlaThrgIuProArgIyPro 20
DB 1 ATGTGGGCTCCACAGAGCTGTGGACAGACATGGAGGCGCACTGAGCCCGCTACCCG 60
QY 21 ProHisSerLeuSerIyProValGlnIleAlaArgThrHisIleAAspValGlyLeuLeu 40
DB 61 CCCACAGGCTTCTTACCAAGAGATGAGATGCCCGAGCGACAGCGAGTGGCTCTG 120
QY 41 GluTrGlnHisIleSerArgAAspIyTrAlaSerHisLeuSerProGlySerIleIleGln 60
DB 121 GAGTACACAGACACACCGCGCATGAGCTCCACCTGCGCGGCTCCATCATTCAG 180
QY 61 ProGlnArgArgProSerIleLeuSerGluPheGlnProGlyIyAAsnIyAAsnSerGln 80
DB 181 CCCACGCGGAGGCGCTCCCTGCTGCTGAGTTCAAGCCCGGAAATGAACGCTCCAG 240
QY 81 GluLeuHisLeuArgProGluSerHisSerIyTrLeuProGluIyIyIySerGluMet 100
DB 241 GAGCTCCACTGGGCGCAGAGTCCCACTATACCTCCGAGCTGGAGAAAGTCAAGATG 300
QY 101 GluPheIleGluSerIyAAspProAlaGlnGluLeuLeuProAAspProLeuLeuAAspPro 120
DB 301 GAGTTCATTGAAGAGAGGCGCTCGGCTAAGAGTCTGCTGAGACCCCTGCTGACCG 360

QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrIysAspArgSer 140
| | | | |
Db 361 TCACCCCTCTGCGCCACGGGCCAGCTGCGGGATCTGAAGACTTCACCAAGAGCCGTAGC 420
| | | | |
QY 141 LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
| | | | |
Db 421 CTGACGGGCAAGCTGGAAACCGGTCTCCCGCCAGCCCCCGACACTGACCTGAGCG 480
| | | | |
QY 161 GluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAsp 180
| | | | |
Db 481 GAGCTGTGCTCCCGCCACGGCTGTCCAAAGAGAGACTATCCAGAAACATGGACCCCGTGGAC 540
| | | | |
QY 181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeu 200
| | | | |
Db 541 CAGAGATACCATGATGAGCAGCAGATCTTAAGCTGAAGAAAGACAGCAACAGCTG 600
| | | | |
QY 201 GJUGJUGJUAAlaLysProProGluProGluLysProValSerProProIleGlu 220
| | | | |
Db 601 GAGGAGGAGGCTGCCAAGCCCGCCGAGCTGAGAAACCCGTGTACCCGCCCGCATCGAG 660
| | | | |
QY 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysIysAlaGluAla 240
| | | | |
Db 661 TCGAAGCACCGCAGCTGTCAGATCATCTACAGAGAAACCGAAAGAGCTGAAGCT 720
| | | | |
QY 241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPro 260
| | | | |
Db 721 GCACATCGGATTCGAGAGCCCTGGGGCCCAAGGTGAGACTGCGCTGTATCAACAGCCC 780
| | | | |
QY 261 SerAspThrArgGlnIleIleIleSerGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
| | | | |
Db 781 TCCGAGACCCGGAGATCATGAGAACTCAAAATTAACCGAGCGATGCGAAGAAAGCTA 840
| | | | |
QY 281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnIleThrLysGlnLysPheCysGln 300
| | | | |
Db 841 ATCTTCTACTTCAAGAGAGAAATCAACCTCGAAACAATGGAGAGAAAGTTCTGCCAG 900
| | | | |
QY 301 ArgTyrAspGlnLeuMetGlnAlaLeuGluLysLysValGluArgIleGluAsnAsnPro 320
| | | | |
Db 901 CCTATGACCACTCATGAGCGCTGGAGAAAGAAAGTGGAGCGCATTCGAGAACACCCC 960
| | | | |
QY 321 ArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340
| | | | |
Db 961 CCGCGCGGGCCAAAGAGAGCAAGGTTCGAGACTACTACAGAAAGAGATTCCTTGAGATC 1020
| | | | |
QY 341 ArgLysGlnArgGluLeuGlnGlnIleLysMetGlnSerArgValGlyGlnArgLysGly 360
| | | | |
Db 1021 CCGAAGCACCGCAGCTGCAGAGAGCCCAATGCAG--AGCGTGGCCAGCGCGGCATGCG 1077
| | | | |
QY 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
| | | | |
Db 1078 CTGTCCATGTCCGCCCGCCGAGCGAGCAAGAGTGTCAAGATCATTCATGCGCTCTCA 1137
| | | | |
QY 381 GJUGJUGJUAAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400
| | | | |
Db 1138 GAGCAGAGAAACTGGAGAAAGCAGATGCGCCAGCTGACCTGATCCGCCCAAGCTGTAC 1197
| | | | |
QY 401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420
| | | | |
Db 1198 GACGCTGACCAAGAGCGCATCAAGTTCACTAACATGAACGGCTTATGGCCGACCCCAAT 1257
| | | | |
QY 421 LysValTyrLysAspArgGlnValMetAsnMetTyrSerGluGlnLysGluThrPhe 440
| | | | |
Db 1258 AAGGTGTACAAAGACCGCAGGTCAATGAACATGTGAGTGAAGAGAGAAAGAGACCTTC 1317
| | | | |
QY 441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
| | | | |
Db 1318 CCGAGAAAGTCAATGAGCATCCCAAGAACTTTGGCTGATGCAATTCCTCGAGAGG 1377
| | | | |
QY 461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480
| | | | |
Db 1378 AAGACAGTGGCTGAGTGGCTCTCTATTACTACTGAAGAAAGATGAAGACTTAAG 1437
| | | | |
QY 481 SerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGln 500
| | | | |

Db 1438 ACCCTGTGAGACGAGGCTATGCGCCCGGCAAGACCCACAGCAACAGCAGCGAG 1497
| | | | |
QY 501 GJUGJUGJUGJUGJUGJUGJUGJUGJUGJUGJUGJUGJUGJUGJUGJUGJUGJUGJUG 520
| | | | |
Db 1498 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1557
| | | | |
QY 521 AspGluLysGluLysGluAlaGluLysGluGluLysGluLysProGluValGluAsn 540
| | | | |
Db 1558 GATGAGAAAGAGAAAGAAAGAGCGGAGAAAGAGAGAAAGCGGAGGTGGAGAAC 1617
| | | | |
QY 541 AspLysGluAspLeuLeuLysGluLysThrAspThrSerGlyGluAspAsnAspGlu 560
| | | | |
Db 1618 GACAAAGAAAGACTCTCTCAAGAGAAAGACAGACCACTCAGGGAGAGAACAGACGAG 1677
| | | | |
QY 561 LysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnIleYArgArgLysGly 580
| | | | |
Db 1678 AAGAGGCTGTGGCTCCAAAGCCCGCAAAACTGCCAAACAGCAGGAGAGACCAAGGC 1737
| | | | |
QY 581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600
| | | | |
Db 1738 CCGATCACCCGCTCAATGGCTAATGAGGCCAACAGAGAGAGCCATCAACCCCGCAGAG 1797
| | | | |
QY 601 SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTyrThrGluGluMet 620
| | | | |
Db 1798 AGCGCGAGCTGGCTCTCCATGAGCTGAATGAGATTCTCGCTGACAGAAAGAAATG 1857
| | | | |
QY 621 GJUTHrAlaLysLysGlyLeuLeuGlnHisGlyArgAsnTyrPheAsnTyrLysLysArgGln 640
| | | | |
Db 1858 GAAACAGCCAAAGAAAGTCTCTTGAAACAGCGCCGCAATGATGCGCATTCGCCCGAGAT 1917
| | | | |
QY 641 ValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGln 660
| | | | |
Db 1918 GTGGGCTCCAAAGACTGTGCGAGTGTAAAGACTTCTTAACACTCAAGAAAGAGGAG 1977
| | | | |
QY 661 AsnLeuAspGluIleLeuGlnGlnHisLysLysLeuLysMetGluLysGluArgAsnAlaArg 680
| | | | |
Db 1978 AACCTCGAGAAATCTTGAGAGCACAAGCTGAAGATGAAGAGAGAGAAACGGCGG 2037
| | | | |
QY 681 ArgLysLysLysLysAlaProAlaAlaAspSerGluGluAlaAlaPheProProValAla 700
| | | | |
Db 2038 AGAAGAAAGAAAGAAACCGCGCGCGCGGCGGCAAGAGAGGCTGCATTCGCCCGCGGTG 2097
| | | | |
QY 701 GluAspGluGluMetGlnAlaSerGlyValSerGlyAsnGluGluLysMetValGluGlu 720
| | | | |
Db 2098 GAGGATGAGAGATGAGAGCGCTCGGGCTGAGACGGGAAATGAGAGAGATGTGGAGAG 2157
| | | | |
QY 721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAla 740
| | | | |
Db 2158 GCTGAA-----GCTC 2166
| | | | |
QY 741 ThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLys 760
| | | | |
Db 2167 ACTGTCAACAACAGCTCAACACACAGAGCATCCCTCTCTCAACACTGAGCGCCAAAG 2226
| | | | |
QY 761 AspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyValaAspGlyProProPro 780
| | | | |
Db 2227 GACACAGGGCAGAAATGGGCCCAAGCCCGACCACTGTGGCGCGAGCGGGCCACCCCA 2286
| | | | |
QY 781 GJYProProThrProPro-ArgArgThrSerArgAlaProIleGluProThrProAlaSe 800
| | | | |
Db 2287 GGGCCACCCACCCACACAGGAGAGACATCCGGCC-CCCACTGAGTCCACCCCGGCTTC 2345
| | | | |
QY 800 rGluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProProPro 820
| | | | |
Db 2346 TGAAGCACCTTAAGCCCTCAAGCCCGCCACACAGACCCCAATTCCTTCACTCTCC 2405
| | | | |
QY 820 oValValProLysGluGluLysGluGluGluThrAlaAlaAlaProProValGluGluG 840
| | | | |
Db 2406 TGTGGTCCCAAGAGAGAGAAAGAGAGAGACCGGACAGCGCCCGCATGTGAGAGAGGG 2465
| | | | |
QY 840 yGluGluGluLysProProAlaAlaGluGluLeuAlaValaAspThrGlyLysAlaGluG 860
| | | | |

Db	2466	GGAGGAGCAGAAAGCCCCCGCGGCTGAGGAGCTGGCAGTGGACACAGGAAAGCCGAGGA	2525
QY	860	uProVal1ySSerSg1uCy8ThrGlUGlUAlAGlUGlUGl1yProAla1ySG1yLYsASP1	880
Db	2526	GCCCGTCAAGAGCGAGTGCACGGAGAAAGCCGAGGAGGGGCGCGCAAGGGCAAGAGCGC	2585
QY	880	AGlUAlaAlAGlUAlaThrAlAGlUGl1yAlaLeu1yAlAGlUy8ySGlUGl1yG1ySe	900
Db	2586	GGAGGCGCCCTGAGGCGCACGGCGCGAGAGGGCGCTCAAGGCACAGAAAGAGAGGGCGGAG	2645
QY	900	rg1yArGLa1aThrThral1ySSerSg1yAlaProGln1ySPSerASPserAlaTh	920
Db	2646	CGGCAAGGCGACACACAGCCAGAGGCTCGAGGCGCCCCCAAGACAGCGCATCCAGTGGCAC	2705
QY	920	rCySSerAlaASPGL1uValASPGLUAlAGlUGl1yG1yASP1yASPana1yGleuLeuSerPr	940
Db	2706	CTGCAGTGCACAGCGAGTGCATGAGCGCGAGGGCGGCGCACAGAAACGGGCTGTCTCCCC	2765
QY	940	oArGProSerLeuLeuThrProThrG1yASPProArGLa1aAna1aSerProGln1yPr	960
Db	2766	AAGGCGCCAGCGCTCTCACCCGAGCTGGGAGACCCCGGCGCAATGCTCACCCCAAGAGCC	2825
QY	960	oLeuASPLeu1ySGl1nLeu1ySGl1nArGLa1aAla1a1eProPro1leglnVal1Thr1y	980
Db	2826	ACTGAGCCTGAAGCGAGCTGAAGCACGAGCGGCTGCCATCCCCCAATCCAGGCTCACCAA	2885
QY	980	sVal1SGl1uPProArGGL1uASPAla1aProThr1ySPPro1aProPro1aAPr	1000
Db	2886	AGTCCATAGCCCCCGGAGAGAGCGAGCTCCACCAAGCAGTCTCCCGACGCC	2945
QY	1000	oProProGlnASPnLeuGln1ProG1uSerASPAlaProGln1nProG1ySerSg1ProAr	1020
Db	2946	GCCACCGCAAAACCTGACGCGGAGAGGAGGAGCGCCCTCAGAGAGCTGGCAGACCCCGG	3005
QY	1020	gG1y1ySSer1yArSg1ProAlaProPro1aASP1ySG1u-----1033	
Db	3006	GGGCAGAGGACAGAGGCCCGGAGACCCCGCGCAGACAGAGGACAGAGAGCCTGTCTT	3065
QY	1034	----AlaPha1a1aAGlUAlAGl1n1yLeuProG1yASPProProCy8ThrThrSg1	1052
Db	3066	CCGAGCTTTCGAGCGAGGCCCGAGAGCTCTGGGGAGACCCCTTGTGAGATTCTCGG	3125
QY	1052	1yLeuProPhaProVal1ProProArGGL1uAl11e1ySP1aSerProH1Sa1aProASP	1072
Db	3126	CCTGGCCTTCCCGCGGCCCGCCCGAGGTATCAAGGCTCTCCCGCATGCCCCGAGCC	3185
QY	1072	oSer1aPhaSer1yAlaProProG1yH1aProLeuProLeuG1yLeuH1aAPThral	1092
Db	3186	CTCAGCCCTTCTCTCAAGCTCCACCTGGTCAACCTGCGCCCTCGGGCTCCATGCATGC	3245
QY	1092	aArGProVal1eupProArGProProThr11eSerASPProProPro1eul1eSerAla	1112
Db	3246	CGGCGCCGCTCGCGCGGCCACCCACCATTTCCAAACCGCGCTCCCTCATCTCTCGC	3305
QY	1112	AlaYH1aSProSerVal1LeuG1uArGGL1n1eG1yAla11eSerGln1yMetSerValG1	1132
Db	3306	CAAGCACCCCAAGCTCTCTCGAGAGGCAATAGTGCCATCTCCCAAGAAATGTGCGTCCA	3365
QY	1132	nLeuH1eVal1Pro1ySerG1uH1a1a1ySa1aProValG1yProVal1ThrMetG1yLe	1152
Db	3366	GCTCCACGTCCTCGTACTCAGAGCATGCCAAAGGCCCGCGGTGGCCCTGTCAACATGGGGCT	3425
QY	1152	uProLeuProMetArSPPro1y8y8LeuAla1aProPhaSerG1yAla1ySGl1nG1n1e	1172
Db	3426	GCCCCCTGCCATGGACCCCAAAAAGCTGGCACCTTCAAGCGAGTGAAGAGGAGCGCT	3485
QY	1172	uSerProArGGL1yGln1aG1yProProG1uSer1eUGl1yAla1ProThralAGln1uAl	1192
Db	3486	GTCGCCAAGGGGCCAGGCTGGGCCACCGAGAGGCTGGGGGTGCCCAACGCCAGAGGCC	3545
QY	1192	aSerVal1eupArGGL1yThralaLeuG1ySerVal1ProG1yG1ySer11eThr1ySG1y11	1212
Db	3546	GTCGCTGCTGAGAGGACAGCTCTGGGCTCAGTTCCGGGCGGAGCATTCACAAAGGCAT	3605

QY	1212	eProSerThrArgValProSerAspSerAlaIleThrTyrArgIysSerIleThrHisGlu	1232
Db	3606	TCCCAGCAGACACGGGATGCCCTCGGACAGCGCCATCATACCGGGCTCCATCACCCACGG	3665
QY	1232	YThrProAlaAspValIleuTyrIysGlyTThrIleThrArgIleIleGlyGluAspSerPr	1255
Db	3666	CACGGCACCTGACGCTCGCTACAGAGGACCATCACAGGATCATCGCGAGAGACAGCC	3725
QY	1255	oSerIrgIeuAspArgGlyArgGluAspSerIeuProIysGlyHisValIleTyrGluGlu	1272
Db	3726	GAGTGGCTTGAACCCCGCGCGGAGAGACAGCTCTGCCAAGGGCCACGTCATCTACGAAG	3785
QY	1272	YValIysGlyHisValIleuSerTyrGluGlyGlyMetSerValThrGlnCyserTyrGlu	1292
Db	3786	CAAGAAAGGGCCACGCTTGTCTTATGAGAGTGGCATGTCTGTGACCCAGTCTCCAAAG	3845
QY	1292	uAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaProIybaThrTyrAs	1312
Db	3846	GACCGGACGAAGCAGCTCAGACCCCGCCATAGACGGCGCCCGCCAAAGCCACCTTAG	3905
QY	1312	pMetMetGluIysArgValGlyArgGluIleSerSerAlaSerIleGlyIleuMetGlu	1332
Db	3906	CATGATGAGAGGGCGCGTGGGAGAGCCATCTCTCGACCGACATCGAAGGTCTCATGG	3965
QY	1332	YArgAlaIleProProGluArgHisSerProHisIleIleuIysGluGlnHisIleAr	1352
Db	3966	CCGTGGCATCCCGCGGAGGCACACAGCCCCCACCACCTCAABAAGACGACCATCCG	4025
QY	1352	gGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnIuAspTyrLeuAr	1372
Db	4026	CGGGTCATACACACAAAGGATCCCTCGGTCTTAGTGAGGCACAGGAGACTACCTGG	4085
QY	1372	GAArgGluAlaIysIleuIeuIysArgGluGlyTThrProProProProProProSerArgAs	1392
Db	4086	TCGGAGGCGCAAGCTCTTAAGCGGAGAGGACGCTCCGCCCCCACCAGCCCTCACGGGA	4145
QY	1392	pLeuThrGlnuAlaTyrIysThrGlnAlaIeuGlyProIeuIysIleuIysProAlaHisGlu	1412
Db	4146	CCTGACCGAGGCTCATAGAGCGAGCGCCCTGGCGCCCTGAAGCTGAAGCCGCGCCATGA	4205
QY	1412	uGlyIleuValAlaThrValIysGluAlaGlyArgSerIleHisGluIleProArgGluGlu	1432
Db	4206	GGGCTGTGGTGGCCACGGTGAAGAGAGCGGCGCTCCATCATGATATCCGCGCGAGGA	4265
QY	1432	uLeuArgHisThrProGluIleuProIeuAlaProArgProIeuIysGluIysSerIleTh	1452
Db	4266	GCTGGGAGACACGCCGAGCTGCGCTCGCGCCCGCGCTCAAGAGGGCTCCATCAC	4325
QY	1452	rGlnGlyThrProIeuIysTyrAspThrGlnAlaSerThrArgIysIleuIysHisAs	1472
Db	4326	GCAAGGACCCCCGCTCAAGTACGACACCGCGCTCCACACTGGCTCCAAAAGACACGA	4385
QY	1472	pValArgSerIleuIleGlySerProGlyArgThrPheProProValHisProIeuAspVa	1492
Db	4386	CGTACGCTCCCTCATCGGACAGCCCGCGGAGCTTCCACCGGTGACCCGCTGATGT	4445
QY	1492	lMetAlaAspAlaArgAlaIeuGluAlaGlyAsTyrGluGluSerIleuIysSerArgPr	1512
Db	4446	GATGGCCGACCCCGCGGACCTGGAACGTGCTCTTAGAGAGAGCCTGAAGACCGCGCC	4505
QY	1512	oGlyThrAlaSerSerSerGlyIysSerIleAlaArgGlyValaProValIleValProGlu	1532
Db	4506	AGGAGCCGCGACAGCTCGGGGGGCTCATTTGGCGCGGGCGCCCGCTCATTTGGCTGA	4565
QY	1532	uIeuGlyIlyeProArgIuSerProIeuThrTyrGluAspHisGlyValaProhealGlu	1552
Db	4566	GCTGGATTAAGCCCGCGGACAGACCCCTGACCTTAGAGACACACGGGACACCTTTGGCGG	4625
QY	1552	yHisIeuProArgGlySerProValThrMetArgGluProThrProArgIeuGlnGluGlu	1572
Db	4626	CCACCTCCACAGAGTTCGCGGTGACCATTCGGAGACCCACGCGCGGCTTCGACGAGGG	4685

QY 1572 ySerLeuSerSerSerIySaIaSeGIuNaSPaRgLyLeuThrSerThProArGIuI 1592
| | | | |
Db 4686 CAGCCTTTGTCAGAGAAAGCATCCAGAGACGAAAGCTGACGCTCGTAGAT 4745
| | | | |
QY 1592 eAlaIySerProHISerThValProGIuNIshISpRoHISProIISerProIyRGI 1612
| | | | |
Db 4746 CGCCAAAGTCCCGCAGACGCGTGCAGACACCAACCCATCTCGCCCTATA 4805
| | | | |
QY 1612 uHISLeuLeuArGIyValISerGIyValaSPLeuTYrARgSerHISILeProLeuAlaH 1632
| | | | |
Db 4806 GCACTCGCTTCGGGGCGTAGAGTGCCTGGAACCTGTATCGAGGCACATCCCTGGCCTT 4865
| | | | |
QY 1632 eASpProThSerIISerProArGIyLISerProLeuSaPaIaIaIaIaIyTYrIleuPr 1652
| | | | |
Db 4866 CCACCCCACTCCATPACCCCGGCGCATCCCTCGAGCAGCGCGCTGCTACTACCTGCC 4925
| | | | |
QY 1652 oArGHSLeuAlaProASnProThTYrProHISLeuTYrProProTYrLeuIISaRGI 1672
| | | | |
Db 4926 CCGAACCTGGGCCCCCAACCCCACTACCCGACCTGTACCACTTACCTCATCCGCG 4985
| | | | |
QY 1672 yTYrProASpThraIaAlaLeuGIuNaSPaRgInThraIISaSPaSPTYrIISerSe 1692
| | | | |
Db 4986 CTAACCCGACAGCGCGCGCTGGAAGAACGGCAGACATCATATGACTACACTACCTC 5045
| | | | |
QY 1692 rGIuNIuMeTHIShISASnThraIaThraIaMeTaIaGIuNaRgIaASpMeIeuaRGI 1712
| | | | |
Db 5046 GCAGCAGATGCACACAAACACGGCGCATGGCCAGCGAGCTGATATGCTGAGGG 5105
| | | | |
QY 1712 yLeuSerProArGIuNIuSerSerIeuaIaLeuANTYrAlaIaGIyProArGIyLISer 1732
| | | | |
Db 5106 CCTCTGCGCCCGGAGACTCTGCTGAGCTCACTACAGCTGCGGGTCCCGAAGCATCAT 5165
| | | | |
QY 1732 eASpLeuSerGIuNIaProHISLeuProValLeuValaProProThProGIyThraProa 1752
| | | | |
Db 5166 CCACTGCTCCCAAGTGCACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5225
| | | | |
QY 1752 aThraIaMeTaSPaRgLeuAlaTYrLeuProThraIaProGIuNIProHISerSerArGI 1772
| | | | |
Db 5226 CACCGCATGAGACGCGCTTGCTACTCCCAACGGGCCCAACCTTCAGCAGCGCCA 5285
| | | | |
QY 1772 sSerSerSerProLeuSerProGIyGIyProThraIISLeuThIyLeProThThraIISe 1792
| | | | |
Db 5286 CAGCAGCTCCCACTCTCCCAAGAGGTCACACACTTGAACAAACCAACCAACCAAGTC 5345
| | | | |
QY 1792 rSerSerGIuArGIuArASpARgASpARgIuArASpARgASpARgIuArGIuIy 1812
| | | | |
Db 5346 CTGCTCCGAGCGGAGAGACCGGGATCGAAGCGGGACCGGGATCGGGAGCGGAAAA 5405
| | | | |
QY 1812 sSerIISLeuThrSerThThraIaGIuNIaIaProIISerIyRgProGIyThrGI 1832
| | | | |
Db 5406 GTCCATCTCTCAGCTCCACCAACGAGGTGAGACGCAACCATCTGAGAACCTGTACAG 5465
| | | | |
QY 1832 uGIuNIserGIySerSerGIySerSerGIyGIyGIyGIyGIySerSerSerArGIProa 1852
| | | | |
Db 5466 GGAGACACAGCGGAGAGCGGAGCGGCGGGGGGGGGGAGCAGACGCGCGCGC 5525
| | | | |
QY 1852 aSerHISerHISaIaHISaGIuNIISerProIISerProArGIuNIaSPaIaLeuGI 1872
| | | | |
Db 5526 CTCCCACTCCCATGCCACCAACGACTGCGCACTCCCTCGAGACCAAGATGCCCTCCA 5585
| | | | |
QY 1872 nGIuNaRgProSerValaLeuHISaSnThraIyMeTySGIyIISerThraIaValGIuPr 1892
| | | | |
Db 5586 GCGAGAACCCAGGTGCTCTACAAACAGGCAATGAAGGATCATCAACCTCTGAGGCC 5645
| | | | |
QY 1892 oSerIySProThValaLeuArGIySerThSerThSerSerProValaArgProAlaIaTh 1912
| | | | |
Db 5646 CAGCAAGACCCACGCTCTGAGGTCACTTCACCTCTCTCAACCCGCTGCCACCTGCAC 5705
| | | | |
QY 1912 rPHeProProAlaThHISaCYSProLeuGIyGIyThThraSPGIyValTYrProThIISe 1932
| | | | |
Db 5706 ATTCCCACTGCAACCACTGCGCACTGGGCGGACCTGATGGGTCTAACCTTACCT 5765
| | | | |
QY 1932 uMeGIuNIProValaLeuLeuProIySGIuAlaProArGIValaIaArgProGIuNaRgProAr 1952
| | | | |

Db 5766 CATGAGCCCGCTTGCTGCGCCCAAGAGGCCCGCGGGTGCCTCGGACAGCGGCGCG 5825
| | | | |
QY 1952 gAlaASpThraGIyHISaIaPHeLeuAlaIySProProAlaArGSeGIyLeuGIuProa 1972
| | | | |
Db 5826 AGCAGACACCGGCACTGCTCTTCTGCGCAAGCCCCAGCCCGCTCGGGCTGAGGCCCG 5885
| | | | |
QY 1972 aSerSerProSerIySGIySerGIuProArGIProLeuValaProProValaSerGIyHISaI 1992
| | | | |
Db 5886 CTCTCTCCCAACCAAGGCTCGAGCCCGGCGCCCTAGGCTCTGCTCTGCGCACCGC 5945
| | | | |
QY 1992 aThraIISaIaRgThraProAlaIySaSPaLeuAlaProHIShISaIaSerProASpPro 2012
| | | | |
Db 5946 CACCATGCGCCGACCCCTGCGCAAGAACTTGCACCTCAACAGCCGACCGGACCGCC 6005
| | | | |
QY 2012 oAlaProProAlaSerAlaSerASpProHISaRGIuIySThGIuNIserIySProHISe 2032
| | | | |
Db 6006 GCGGCACTGCTCGGCTCGGCTCGGACCGGACCGGAAAAAGATCAAAAGTAAACCTTTTC 6065
| | | | |
QY 2032 rIISerGIuLeuGIuLeuArGIySerIeGIyTYrHISGIySerSerTYrSerProGIuGI 2052
| | | | |
Db 6066 CATCCAGGAACCTGAACTCGCTTCTGAGTTACACGCGCAGACGCTACAGCCCGCAAG 6125
| | | | |
QY 2052 yValGIuNIProValaSerProValaSerSerProSerLeuThraIISaSPaIySGIyLeuProIy 2072
| | | | |
Db 6126 GGTGAGCCCGTTCAGCGCTGTGACTCACTGACCTGACCCAGACAGAGAGGCTCCCA 6185
| | | | |
QY 2072 SHISLeuGIuGIuLeuASpIySerHISLeuGIuGIyLeuLeuArGIyProGIuProGI 2092
| | | | |
Db 6186 GCACTGGAAGACTGACAAAGACCACTGAGAGGGAGCTGCGGCCAAGACGCAAG 6245
| | | | |
QY 2092 yProValaIyLeuGIyGIyGIuAlaIaHISLeuProHISLeuARGIyProLeuProIySe 2112
| | | | |
Db 6246 CCCCCTGAAGCTTGGGGGGAGGCGGCCCACTCCACACTTGGCGCCCTGCTGAGAG 6305
| | | | |
QY 2112 rGIuNIProSerSerSerProLeuLeuGIuThraIaProGIyValaIySGIyHISaIaRgVa 2132
| | | | |
Db 6306 CAGCGCTGCTCAGCGCTGCTCTCCAGACCGCGCGGAGGTCAAGGTACACAGGGGT 6365
| | | | |
QY 2132 IValaThraIaIaGIuNIHISerGIuValaIISerGIuNAAPTYrThraGIHISaIaR 2152
| | | | |
Db 6366 GGTCACTCGGCGGACCAATCATGAGGTATCATACAGACTACACCGGACCAACCG 6425
| | | | |
QY 2152 oGIuNIuLeuSerAlaProLeuProAlaProLeuTYrSerPHeProGIyAlaSerCySP 2172
| | | | |
Db 6426 ACAGCAGCTCAAGGACCCCTGCGCCGCTCACTCTCTCTGCGGGCCAGCTGCGCC 6485
| | | | |
QY 2172 oValaLeuASpLeuArGIyProProSerASpLeuTYrIISerProProProASpHISaIa 2192
| | | | |
Db 6486 CGTCTGGACTTCGCGCGCCCAACCGTACCTTACCTCCGCGCCCGGACCATGTGC 6545
| | | | |
QY 2192 aProAlaArGIySerProHISerGIuGIyIySaArGSeRProGIuNIProASpIySTH 2212
| | | | |
Db 6546 CCGGCGCTGCTGCCCGCCCAAGCAAGGAGGCAAGAGTCTCCAGACCAAAACAAGAC 6605
| | | | |
QY 2212 rSerValaLeuGIyGIyGIuNaSPGIyLISerProValaSerProProGIuNIyMeTh 2232
| | | | |
Db 6606 CTGCGTCTGGGGGTGAGAGACGGTATTGAACCTGTGTCCCAACCGAGGGCATGAG 6665
| | | | |
QY 2232 rGIuNIProGIyHISerArGIySerAlaIaIyTYrProLeuLeuTYrARgASpGIyGIuNIuH 2252
| | | | |
Db 6666 GAGGCCAGGACATCCCGAGGTGTGTATCCGCTGCTGACCGGAGTGGGAAACAAGAC 6725
| | | | |
QY 2252 rGIuNIProArGIySerIySerProGIyASnThSerGIuNIProProAlaPHePh 2272
| | | | |
Db 6726 GAGGCCAGAGATGGCTTCAAGTCTCAAGGCAACAAGCAGACCGGCAACCTTTCT 6785
| | | | |
QY 2272 eSerIyLeuThraGIuNIserASnSerAlaMeTaIySerIySGIyGIuNIleASpIy 2292
| | | | |
Db 6786 CAGCAAGCTGACGAGAGAACTCCGCTATGTCAAGTCCAAGAAACAGAGATCAACAA 6845
| | | | |
QY 2292 eIyLeuASnThraIISaSPaRgANCIuProGIuTYrASpNIISerGIuNIyThraGI 2312
| | | | |

Db 6846 GAAGCTGACACCCACCAACCGGAATGAGCTGAAATACATATCAGCAGCCTGGAGCGA 6905
Qy 2312 uilepheAmePProAlaIleThrGlyLeuMetThrTyArgSerGlnAlaVa 2332
Db 6906 GATCTTCAATATGCCCCCATCACCAGAACGGCCTTATGACTTATGAAAGCCAGGGCGT 6965
Qy 2332 lGInGUHlAAlaSerThranMetGlyLeuGlnAlaIleIleArgValaLeuMetG 2352
Db 6966 GCAGGACATGCGACCAACATGGGGCTGGAGCCATATTTGAAAGGACCATGGG 7025
Qy 2352 ylyeTyArgGlnITPGLuIuSerProProLeuSerAlaAmaLaApeApeProLeuAs 2372
Db 7026 TAATATAGACAGGAGGAGAGTCCCGCGCTCAGGCGCCATCTTTAACTCTGAA 7085
Qy 2372 nAlaSerAlaSerleuProAlaAlaMetProIleThrAlaAlaApeGlyArgSerApeH 2392
Db 7086 TGCAGTGCAGCCTGCCCGCTGTATGCTTATACCGCTGCTGACGAGCGAGTGAACA 7145
Qy 2392 sThrLeuThrSerProGlyGlyGlyGlyValaValaSerGlyArgProSerSerAr 2412
Db 7146 CACACTCAGCTCGCCAGGTGGCGGGAGGAGCCAGAGTCTTGAGACCCAGACCG 7205
Qy 2412 GlyValaIySerProAlaProGlyLeuAlaSerGlyAAspArgProProSerValSerSe 2432
Db 7206 AAAAGCCAGATCCCGCGCCCTGGGCTGGCATCTGGGAGACCGGACCTCTGTCTCTC 7265
Qy 2432 rValHisSerGlnGlyAAspCyAAnaArgThrProLeuThraApeValITPGLuAs 2452
Db 7266 AGTGACATCGAGAGGAGACTGCAACCGCCGACCGCTCAACCAACCGCGTGGAGGA 7325
Qy 2452 pATGProSerSerAlaGlySerThrProPheProTyArgProLeuIleMetArgLeuG 2472
Db 7326 CAGGCGCTCGCCGAGGTTCCAGGCATTCCTTACAACTCCCTGATCATGGGCTGCA 7385
Qy 2472 nAlaGlyValaMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeuAl 2492
Db 7386 GGGGGGTGTATGGCTTCCACCCACCGGGGCTCCCGGGGCGAGCGGGCCCTCGC 7445
Qy 2492 aglyProHisAlaIleThraApeGlnIuProLyPProLeuLeuCySerGlnTyArgIuTh 2512
Db 7446 TGGCCCCCAGCAGCCTGGGAGAGAGAGCCCAAGCCACTGCTGTGCGATGACGAGAC 7505
Qy 2512 rLeuSerAAspSerGlu 2517
Db 7506 ACTCTCGAGCAGCGAG 7521

RESULT 9
ADL13812
ID ADL13812 standard; DNA: 7524 BP.
XX
AC ADL13812:
XX
DT 06-MAY-2004 (first entry)
XX
DE Osteoarthritis-associated polymorphic nucleotide #344.
XX
KM ds: gene: osteopathic; antiinflammatory; antiarthritic; gene therapy;
KM joint space narrowing; osteophyte development; joint pain;
KM osteoarthritis; SNP; single nucleotide polymorphism.
XX
OS Homo sapiens.
XX
PN MO2003054166-A2.
XX
PD 03-JUL-2003.
XX
PF 19-DEC-2002; 2002MO-US041225.
XX
PR 20-DEC-2001; 2001US-0342603P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Jones KA, Schaffer A;

XX
DR WPI: 2003-559141/52.
XX
PT Determining susceptibility of an individual to joint space narrowing,
PT osteophyte development and/or joint pain comprises identifying whether
PT the individual has at least one polymorphism in a polynucleotide encoding
PT a protein.
XX
PS Disclosure; SEQ ID NO 344; 297pp; English.
XX
CC The invention relates to a method of determining susceptibility of an
CC individual to joint space narrowing and/or osteophyte development and/or
CC joint pain comprising identifying whether the individual has at least one
CC polymorphism in a polynucleotide encoding at least one of the protein
CC listed in the specification. The methods, composition and agent are
CC useful for modulating the susceptibility of an individual to joint space
CC narrowing and/or osteophyte development and/or joint pain that is
CC associated with a disease, preferably osteoarthritis. The cell line and
CC the non-human animal are useful for screening for an agent for diagnosing
CC an individual having susceptibility to joint space narrowing and/or
CC osteophyte development and/or joint pain. This sequence corresponds to
CC the polynucleotide encoding a protein listed in the specification. (Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences).
XX
SQ Sequence 7524 BP; 1636 A; 2728 C; 2213 G; 947 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 0 Length: 7524
Score: 12978.00 Matches: 2487
Percent Similarity: 98.57% Conservative: 3
Best Local Similarity: 98.46% Mismatches: 9
Query Match: 98.21% Indels: 28
DB: Gaps: 3
US-09-522-753-5 (1-2517) x ADL13812 (1-7524)
Qy 1 MetSerGlySerThrGlnLeuValAlaGlnThrThraApeValITPGLuProArgTyPro 20
Db 1 ATGTGGGCTCCACACAGCTGTGGCACAGCTGGAGGGCCACTGAGCCCGCTACCG 60
Qy 21 ProHisSerleuSerTyProValGlnIleAlaArgThrHisThraApeValGlyLeuLeu 40
Db 61 CCCCACAGCTTTCCTTACCAAGTGCAGATGCCCCGAGACGACGATCGGGCTCTCTG 120
Qy 41 GluTyGlnHisHisSerArgApeTyAlaSerHisleuSerProGlySerIleIleGln 60
Db 121 GAGTACACAGACCACTCCGCGACTATGCTCCCACTGTGCGCCGCTCATCATCAG 180
Qy 61 ProGlnArgArgArgProSerleuLeuSerGlnPheGlnProGlyLeuArgSerGln 80
Db 181 CCCACGGCGGAGGCGCTTCCCTGCTGTGAGTTCAGAGCCCGGAAATGAACGTCAG 240
Qy 81 GluLeuHisleuArgProGlnSerHisSerTyLeuProGlnleuGlyLeuSerGlnMet 100
Db 241 GAGCTCCAGCTGGGCGCAAGTCCACTCACTACTCCCAAGCTGGGAGTCAAGATG 300
Qy 101 GluPheIleGlnSerIyArgProArgLeuGlnLeuLeuProApeProleuLeuArgPro 120
Db 301 GAGTTCATTGAAGAAGCAGCGCTCGGCTAGAGACTGCTGACCCCTGCTGCGACCG 360
Qy 121 SerProleuLeuAlaThrGlyGlnProAlaGlySerGlnAAspLeuThrIyAAspArgSer 140
Db 361 TCACCCCTGCTGCGCACGGGCGAGCTGGCGGATCTGAAGACTCAACCAAGACCGTAGC 420
Qy 141 LeuThrGlyLyLeuGlnProValSerProProSerProProHisThraApeProGlnLeu 160
Db 421 CTGACGGGAGAGTGAACCGGTGTCTCCCGCAGGCCCGCCGACACTGACCTTAGCTG 480
Qy 161 GluLeuValProProArgLeuSerIyGlnGlnLeuIleGlnAAspApeValAAsp 180
Db 481 GAGCTGTGCTCGCCACAGGCTGTCTCAAGAGAGAGTATCCAGAACATGACCGCGTGGAC 540

QY 181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnIleu 200
 Db 541 CGAGAGATACCATGTAGTAGAGCAGACGATCTTAAGCTCAGAGAGAGACGACGCTG 600
 QY 201 GUGUGUUAUAAUlybProGluProGluLysProValSerProProPoliGlu 220
 Db 601 GAGGAGGAGGCTGCCAAGCCGCCAGCCTGAGAGCCGCTGTCAACCGCCGCATCGAG 660
 QY 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysLysLys 240
 Db 661 TCGAAGCACCGCAGCCTGTGTGATCATCTACAGAGAACCGGAGAGAGGCTGAAGCT 720
 QY 241 AlaHisArgLysLeuGluGluLysLeuGluProGlnValGluLeuProLeuTyrAsnGlnPro 260
 Db 721 GCACATCGGATCTGGAAAGCCTGGGGGCCCAAGTGAAGCTGCGCTGTACAAACGAGCC 780
 QY 261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
 Db 781 TCCGACACCCGGCAGATCATGAAACATCAAAATTAACAGCGCATGCGGAGAAAGCTA 840
 QY 281 IleLeuTyrPheLysArgArgArgAsnHisAlaArgLysGlnThrLysGlnLysPheCysGln 300
 Db 841 ATCTGTACTTCAAGAGAGAAATCAAGCTCGAACAATGGAGCAGAAATTCTGCCAG 900
 QY 301 ArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluLysLysLysLysLys 320
 Db 901 CGCTATGACCACTCATGAGAGGCTGGAGAAAGAAAGTGAAGCGCATTCAGAAACAAACCC 960
 QY 321 ArgArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340
 Db 961 CGCGCGCGCGCCAGAGAGAGCAAGGTTCCGAGTACTACAGAAAGCAAGTTCCCTGAGATC 1020
 QY 341 ArgLysGlnArgGluLeuGlnGlnIleuArgMetGlnSerArgValGluLysGlnArgLysGly 360
 Db 1021 CCGAAGCACCGGAGCTTGCAAGAGCCCATGCAAG--AGGGTGGCCAGCGCGGCAAGTGG 1077
 QY 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
 Db 1078 CTGTCCATGTCCCGCCCGCAGCGAGCAAGAGTGTCAAGATCATTCATGAGCTCTCA 1137
 QY 381 GUGUGUUAUAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProPheMetLeuTyr 400
 Db 1138 GAGCAGAGAAACCTGAGAGAGAGATGCCAGCTGGCGTGTATCCCGCCCATGTGCTATC 1197
 QY 401 AspAlaAspGlnGluArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspPromet 420
 Db 1198 GACGCTGACCAAGCGCATCAAGTTCACTACATACATAGCGGCTTATGGCCCAACCCCATG 1257
 QY 421 LysValIleLysAspArgGlnValMetAsnMetTrpSerGluGlnLysLysLysLysLys 440
 Db 1258 AAGGTGTACAAAGACCGCCAGGTCATGAACATGTGAGTGAAGCAGAGAGAGACCTTC 1317
 QY 441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
 Db 1318 CCGGAGAAATTCATGACAGATCCCAAGAACTTTGGCTGATCGCATCATTTCCCTGAGAG 1377
 QY 461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480
 Db 1378 AAGACAGTGGCTGAGTGGCTCTCTATTTACTTACTTACCTAAGAAATGAGAACTTAAG 1437
 QY 481 SerLeuValArgSerTyrArgArgArgLysLysSerGlnGlnGlnGlnGlnGlnGln 500
 Db 1438 AGCTGTGTAGAGAGAGGCTATCGGCGCCGCGGCAAGCAGCAGCAACAAACGAGCAG 1497
 QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
 Db 1498 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1557
 QY 521 AspGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLys 540
 Db 1558 GATTGAAAGAGAAAGAAAGAGGCGGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617

QY 541 AspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAsnAspGlu 560
 Db 1618 GACAAGAGAAAGCTCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1677
 QY 561 LysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGly 580
 Db 1678 AAGGAGGCTGTGGCTTCCAAAGCCGCAAAATTCGCAACAGCAGAGAGAGAGAGAGAG 1737
 QY 581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600
 Db 1738 CGCATCACCCGCTCAATGGCTTAATGAGGCCAAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 1797
 QY 601 SerAlaGluLeuAlaSerMetGluLeuAsnGlnLysSerArgTyrThrGluGluLysMet 620
 Db 1798 AGCGCGAGCTGGCTCTCCATGAGCTGAATGAGATTTCTCGTGAACAAAGAGAGAGAG 1857
 QY 621 GlnThrAlaLysLysGluLeuLeuGlnHisGlyArgAsnTrpSerAlaIleAlaArgMet 640
 Db 1858 GAAACAGCCAGAGAAAGGTCTCTGGAACACGAGCCGCACTGGTCCGCGCATCGCCGAGATG 1917
 QY 641 ValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGln 660
 Db 1918 GTGGGCTCCAAGACTGTGTGCGAGGTAAAGACTTCTACTCACTACAAAGAGAGAGAG 1977
 QY 661 AsnLeuAspGluIleLeuGlnGlnHisLysLysLeuLysMetGluLysGluLysArgAsnAlaArg 680
 Db 1978 AACCTCGATGAGATCTTGCAGCAGACAAAGTGAAGATGGAAGAGAGAGAGAGAGAGAG 2037
 QY 681 ArgLysLysLysLysAlaProAlaAlaSerGlnGluGluAlaAlaPheProProValVal 700
 Db 2038 AAGAAAGAAAGAAAGAGCCCGCGCGCGCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 2097
 QY 701 GlnAspGlnGluMetGluAlaSerGlyValSerGlyAsnGluGluLysMetValGluGlu 720
 Db 2098 GAGATGAGAGATGAGAGAGCGTCCGGCGTGAAGGAAATGAGAGAGAGATGCTGAGAGAG 2157
 QY 721 AlaGluAlaLeuHisAlaSerLysAsnGluValProArgGlyGluCysSerGlyProAla 740
 Db 2158 GCTGAA-----GCTC 2166
 QY 741 ThrValAsnAsnSerSerAspThrGlnSerIleProSerProHisThrGluAlaAlaLys 760
 Db 2167 ACTGTCAACACAGCTCAGACACCGAGAGCATCTCTCTCTCACTAAGAGCCGCGCAAG 2226
 QY 761 AspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProPro 780
 Db 2227 GACACAGGGCAGAAATGGGCCCAAGCCCGCAGCACCTGGCGGCGGAGAGAGAGAGAG 2286
 QY 781 GlyProProThrProPro-ArgArgThrSerArgAlaProIleGluProThrProAlaSer 800
 Db 2287 GGGCCACCCACCCCAACCCGAGAGACATCCCGGCC-CCCACTGAGTCCACCCCGGCTTC 2345
 QY 800 rGluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProProPro 820
 Db 2346 TGAAGCACCTTAAGCCCTCAGCCCGCCACAGACACCCCATTTCCCTTTCACCTCTCC 2405
 QY 820 oValValProLysGluGluLysGluGluThrAlaAlaAlaProProValGluGluGlu 840
 Db 2406 TGTGGTCCCAAG 2465
 QY 840 YGluGluGluLysProProAlaAlaGluGluLeuAlaValAlaAspThrGlyLysAlaGluGlu 860
 Db 2466 GAG 2525
 QY 860 uProValLysSerGluCysThrGlnGluAlaGluGluGluGluProAlaLysGlyLysAspAl 880
 Db 2526 GCCCGTCAAG 2585
 QY 880 agluAlaAlaGluAlaThrAlaGluGluValAlaLeuLysAlaGluLysLysGluGluGly 900
 Db 2586 GGAAGCGCGTGAAG 2645
 QY 900 rGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThr 920

Dh	2646	CGGCGAGGGCCACACACACCGCAGAGCTCGGGGCGCCCCCGACGAGCACGACTCCAGTGGCAC	2705
Oy	920	TCySerAlaAspGluValAspGluValAGLuGlyGlyAspIysAsnAlaGluLeuLeuSerPr	940
Dh	2706	CTGCAGTGCAGACGAGGTGAGTGAAGCCGAGGGCGCGCACAAAGAACCGGCTGTGTCTCCC	2765
Oy	940	QArGPProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysPr	960
Dh	2766	AAGGCCACGAGCTCTCACCCGACCTGGGGAGACCCCGGGGCGCAATGCTCTCACCCCGAAGAGCC	2825
Oy	960	OLEuAspIleuLysGlnLeuLysGlnIleArgAlaAlaIleProProIleGlnValThrIly	980
Dh	2826	ACTGGAACCTGAAGCGCTGAAGACAGCGAGCGAGCTGCATCTCCCGCCACATCCAGGTCAACCA	2885
Oy	980	AVAlHisGluProProArgGluAspAlaAlaProThrIysAspAlaProProAlaProPr	1000
Dh	2886	AGTCATATAGCCCCCGGGGAGAGCGAGCTCCACCAAGCAGACTCCCGCCAGCCCCAC	2945
Oy	1000	OProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProAr	1020
Dh	2946	GCCACCGCAAAACCTGACGCGGAGAGGAGCGCCCTCAGCAGCTGGCAGACAGCCCCCG	3005
Oy	1020	GGIlyAsSerIArgSerProAlaProProAlaAspIlyGlu-----	1033
Dh	3006	GGGCAAGAGCGAGAGCCCGGAGACCCCGCGACAGAGGCGAGAGAGCCTGTGTTCTT	3065
Oy	1034	----AlaPheAlaIaGluValAGIlyLysLeuProGlyAspProProCysTrpThrSerG1	1052
Dh	3066	CCGACGCTTCGACGCGAGGCCAGAGACTCGTGGGAGACCCCTTGCTGAGCTTCGCG	3125
Oy	1052	YLLeuProPheProValProProArgGluValIleLysAlaSerProHisAlaProAspPr	1072
Dh	3126	CCTGGCCTTCCCCCGTGGCCCCCGGAGGTATCAAGGCTCTCCCGATGCCCCGAGACC	3185
Oy	1072	OSerAlaPheSerItyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAl	1092
Dh	3186	CTCAGCCTTCTCTTAACGCTCCACCTGGTCAACCACTGCCCTGGGGCTCCATGACATGCG	3245
Oy	1092	AAArgProAlaLeuProArgProProThrIleSerAsnProProProLeuIleSerSerAl	1112
Dh	3246	CCGGCGCGCTCTGGCGGCGGCCACCCACCATCTCCAAACCGGCTCTCCATCTCTCTGCG	3305
Oy	1112	ALyHisIAsProSerValIleuGlnIArgGlnIleGlyAlaIleSerGlnGlyMetSerValG1	1132
Dh	3306	CAAGCACCCCAAGCGCTCTCGAGAGGCAATATGCTGCATCTCCAAAGGATGTGCTGCCA	3365
Oy	1132	nLEuHisValProTyrSerGlnHisAlaIlyAsAlaProValGlyProValThrMetGlyLe	1152
Dh	3366	GCTCCAGCGTCCCGTACTCAGAGCATGTCMAAGGCCCGGTGGGCTGTGCACATGGGGCT	3425
Oy	1152	uProLeuProMetAspProIlyIlyLysLeuAlaProPheSerGlyValIlyGlnGlnIle	1172
Dh	3426	GCCCCCTGCCCATGAGCCCAAAAGACTGGGCACTTCAAGCGAGTGAAGCAGGAGCGCT	3485
Oy	1172	uSerProArgGlyGlnIaGlyProProGlnIleSerLeuGlyValaProThrIaGlnGlnAl	1192
Dh	3486	GTCCCCCAGGGGCGAGGCTGGGGCCACCGGAGAGCTGGGGGTGGCCACAGCCACGGAGGC	3545
Oy	1192	ASerValIleuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrIlyGlyI1	1212
Dh	3546	GTCGGTGCTGAGAGGAGCAGCTCTGGGGCTCATCTCCGGGCGGAGAGCATCAACAAAGCAT	3605
Oy	1212	eProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHisG1	1232
Dh	3606	TCCACAGACACAGGGGTGCCCTCGGACAGGCGCATCAATACCGCGCTCCATCAACCCACGG	3665
Oy	1232	YThrProAlaAspValIleuTyrLysGlyThrIleThrArgIleIleGlyGlnIuAspSerPr	1252
Dh	3666	CAGCCGACGACTACGTCTCTGTCMAAGGGCACCATCAACGAGATCATCGGCGAGGACACCCC	3725
Oy	1252	OSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrGlnG1	1272

Dh	3726	GAAGTCGCTTGGACCGCGGGCCGGAGAGACAGCTCGCCCAAGGGCCACGTCATCTACGAAGG	3785
Qy	1272	YLysArgSGlyHisValLeuSerTYrGluGlyGlyMetSerValThrgIncySerIysGly	1292
Dh	3786	CAAGAAAGGGCCACGTCCTGTCCTATGAGAGTGGCATGTCTGTGACCCAGTGTCTCAAGGA	3845
Qy	1282	uAspGlyArgSerSerSerGlyProProHisGluThrAlaIleProIlySerGlyThrAs	1312
Dh	3846	GGACGGCGAAGACAGCTCAGAGACCCCCCAAGACGGCCGCCCCCAAGCCACCTTAGGA	3909
Qy	1312	pMetMetCyluGlyArgValGlyArgAlaIleSerSerIleSerIleGlyIleuMetGly	1332
Dh	3906	CATGTATGAGAGGACCGCGTGGGCAAGGCCATCTCTCAGCCAGCATTCGAAGGCTTCATGGG	3965
Qy	1332	YArgAlaIleProProGluArgHisSerProHisIleuLeuysGluGlnHisIleLeu	1352
Dh	3966	CCGTGCCATCTCCGCCGGAGCCACACAGCCCCCACACTCTCAAAAGACGACCACTCCG	4025
Qy	1352	gGlySerIleThrgInGlyIleProArgSerTYrValGluAlaGlnGluAspTYrLeuAr	1372
Dh	4026	CGGGTCATCAACAAGAGGATCCCTCGGTCTTAGTGAAGCACAGGAGGACTACCTGGC	4085
Qy	1372	gArgGluAlaIysLeuLeuLeuysArgGlyGlyTYrThrProProProProProProSerArgAs	1392
Dh	4086	TCGGAGAGGCCAAGCTCTCTAAAGCGGAGGAGCACGCTCCGCCCCACCGCTCTACGGGA	4145
Qy	1392	pLeuThrgIuAlaTYrIysThrgInAlaLeuGlyProLeuysIleuysAspProAlaHisGly	1412
Dh	4146	CTTAGCCAGGCGCTTACAGACGACGAGCCCTGGGCCCTTGAAGCTTGAAGCCGGCCATGA	4205
Qy	1412	uGlyLeuValAlaThrValIysGluAlaGlyArgSerIleHisGluIleProArgGluGly	1432
Dh	4206	GAGCGGTGGTGCACAGGTGAAGAGAGGGGGCGCTTCATCATATGATCCCGCGCAGGA	4265
Qy	1432	uLeuArgHisThrProGluLeuProLeuAlaProArgProLeuysGlyGlySerIleTh	1452
Dh	4266	GCTCGGGCACACGCCCGAGCTGCCCTGGCCCCCGCGCTCTCAAGGAGGGCTCCATCATC	4325
Qy	1452	rGlnGlyThrProLeuysTYrAspThrArgAlaSerThThrgIysSerIlyHisAs	1472
Dh	4326	GCAAGGACCCCCGCTCAAGTACGACACGGGGCGCTCCACCACTGGCTTCAAAAAGACGA	4385
Qy	1472	pValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspVa	1492
Dh	4386	CGTAGCGCTCCATTATCGGAGGCCCGCGGCCGAGCTTCCACCGGTGACCCGCTGAGTGT	4445
Qy	1492	IleMetAlaAspAlaArgAlaLeuGluArgAlaCysTYrGluGlnSerIleuysSerArgPr	1512
Dh	4446	GATGGCCGACCGCCCGGACACTGGAACTGGCTCTTAGAGAGAGCCTGAAAGACCCGGCC	4505
Qy	1512	oGlyTYrHisAserSerSerGlyGlySerIleAlaArgGlyValaProValIleValaProGly	1532
Dh	4506	AGGAGCCGCCACGAGCTCGGGGGGCTCATTTGGCGGGGGCGCCCGGTGCAATTGGCTGA	4565
Qy	1532	uLeuGlyLysProArgGlnSerProLeuThrTYrGluAspHisGlyValaProPheAlaGly	1552
Dh	4566	GCTGGGTAAAGCCCGGAGAGACCCCCCTGACCTTAGAGACCAAGGGGACACCTTTGCCGG	4625
Qy	1552	yHisLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGly	1572
Dh	4626	CCACCTCCCAAGAGTTGCGCCCGTACCATTCGGGAGACCCACGCGCGGCTGACGAGAGGG	4685
Qy	1572	ySerLeuSerSerSerIlyHisAserGlnAspArgIlyHisLeuThSerThrProArgGlyIle	1592
Dh	4686	CAGCTTTTCGTCACGAAGGCATCTCCAGGACCGAAMACTACGTCAGCGCTGTGAGAT	4745
Qy	1592	eAlaIysSerProHisSerThrValProGlnHisArgProHisArgIleSerProTYrGly	1612
Dh	4746	CGCCAAATCCCCGACACGACCGTGGCCCGACACCAACCCCACTCTCGCCCTTAGGA	4805
Qy	1612	uHisLeuLeuArgGlyValSerGlyValaAspLeuTYrArgSerHisIleProLeuAlaPh	1632
Dh	4806	GCACCTCTCTCGGGGCTGAGTGGCGGTGACCTGTATCGACGCACTCCCTCGGGCTT	4865

[illegible][illegible]

Db 7026 TAAATATACCAAGTCGGAAGAGTCCCGCCGCTCAGCCCAATGCTTTAAACCTCTGAA 7085
Qy 2312 nAlSerAlaSerLeuProAlaAlaMetProLeuThrAlaAlaSerGlyAAsSerAAsP 2392
Db 7086 TCCAGTGTCCAGCCCTGCGGCTGTATCCCTAAACCCCTGTGACGAGAGAGCA 7145
Qy 2392 sThrLeuThrSerProGlyGlyGlyGlyAlaValSerGlyAAsProSerSerA 2412
Db 7146 CACACTACCTCTGCGAGTGGCGGCGGAGGAGCCAGCTCTTGGCAGACCCAGCAGCCG 7205
Qy 2412 GlyAlaAlaSerSerProAlaProGlyLeuAlaSerGlyAAsPProProSerValSer 2432
Db 7206 AAAAGCCAAATGCCCGGCGGCTGGCATCTGGGAGACCGGACCCCTCTGCTCCTC 7265
Qy 2432 rValHisSerGlyGlyAAsPProGlyAAsPProGlyAAsPProGlyAAsPProGlyAAs 2452
Db 7266 AGTCACACTCGGAGAGAGACTCAACCGCGGACCGCTCACCAACCGCTGTGGAGAG 7335
Qy 2452 PArgProSerSerAlaGlySerThrProPheProTyraProLeuIleMetArgLeuG 2472
Db 7326 CAGGCCCTCGTCCGAGGTTCCAGCCATTCCTTCAACCCCTGATCATGGGCTGCA 7385
Qy 2472 nAlGlyValMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeuAl 2492
Db 7386 GCGGGGTGTGATGCTTCCACCCGCGGCTCCCGGCGGAGCGGCGGCGGCGGCTCCG 7445
Qy 2492 aGlyProHisHisAlaThrPheProGlyLeuProLeuLeuGlySerGlyProGlyLeu 2512
Db 7446 TGGGCCCCACCAAGCTGTGGAGAGAGCCAGACCACTGCTGTGCTGAGTACGAGAC 7505
Qy 2512 rLeuSerAAsPProGly 2517
Db 7506 ACTCTCCGACAGCCAG 7521
RESULT 10
ID ACA62249 standard; cDNA; 8686 BP.
XX ACA62249;
XX AC ACA62249;
XX 12-AUG-2003 (first entry)
DE cDNA encoding human nuclear receptor corepressor SMRte.
XX Human; sex; gene; SMRte; nuclear receptor corepressor; gene therapy;
XX tissue typing; cancer.
XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT CDS 157..7680
FT /tag= a
FT /product= "SMRte"
XX US2003027137-A1.
XX
XX 06-FEB-2003.
XX
XX 27-MAR-2001; 2001US-00819104.
XX
XX 29-MAR-2000; 2000US-0193138P.
XX
XX (CHEN/) CHEN J D.
XX
XX Chen JD;
XX
XX WPI; 2003-466139/44.
XX
XX P-PSDB; AB061812.
XX
XX New SMRte proteins and nucleic acids, useful in gene therapy, predictive
PT medicine, therapeutic or prophylactic treatment, chromosome mapping,
PT tissue typing and in forensic biology.

XX
PS Claim 2; Page 32-41; 90pp; English.
XX
CC The invention relates to an isolated SMRte nucleic acid molecule. The
CC nucleic acids are useful in gene therapy, as hybridization probes for
CC identifying SMRte-encoding nucleic acid molecules and as primers for
CC amplifying of SMRte nucleic acid molecules. The polypeptides are useful
CC as immunogens to raise anti-SMRte antibodies. The SMRte molecules are
CC useful as targets for discovering and developing modulating agents to
CC regulate a variety of cellular processes, in screening assays, in
CC predictive medicine, in therapeutic or prophylactic treatment, in
CC chromosome mapping, tissue typing and in forensic identification of a
CC biological sample. Modulators of SMRte are useful for treating or
CC preventing a condition associated with aberrant SMRte protein or nucleic
CC acid expression or activity, such as cancer. The present sequence
CC represents cDNA encoding the human nuclear receptor corepressor SMRte
XX
XX Sequence 8686 BP; 1881 A; 3069 C; 2562 G; 1174 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 0 Length: 8686
Score: 12978.00 Matches: 2487
Percent Similarity: 98.57% Conservative: 3
Best Local Similarity: 98.46% Mismatches: 9
Query Match: 98.21% Indels: 28
DB: 8 Gaps: 3
US-09-522-753-5 (1-2517) x ACA62249 (1-8686)
Qy 1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTrpPro 20
Db 157 ATGTGGGGCTCCACAGCCGCTGGGACAGACGCGAGGGCCACTGAGCCCGCTTACCG 216
Qy 21 ProHisSerLeuSerTrpProValGlnIleAlaArgThrHisThrAAsPValGlyLeuLeu 40
Db 217 CCCACAGCCTTCTTCAACCAAGTGCAGATGCGCCGAGACAGACAGCGTGGGCTCCTG 276
Qy 41 GluTrpGlnHisHisSerArgAAsPTrpAlaSerHisLeuSerProGlySerIleIleGln 60
Db 277 GAGTACCAAGACCACTCCCGGACTATGCTCCCACTGTGCGCCGCTCATATATCAG 336
Qy 61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAAsGlnArgSerGln 80
Db 337 CCCAGCGGGAGAGGCCCTCCGCTGTGTGAGTTCAGCCGCGGAATGAACGGTCCAG 396
Qy 81 GluLeuHisLeuArgProGluSerHisSerTrpLeuProGluLeuGlyLeuSerGluMet 100
Db 397 GAGCTCACCTGCGGCGCAGAGTCCCACTATACCTGCCAGCGTGGGAAGTCAGAGATG 456
Qy 101 GluPheIleGluSerIleArgProArgLeuGluLeuProAAsPProLeuLeuArgPro 120
Db 457 GAGTTCATTAAAGCAAGCCCTCGGCTAGACTGTGCTGACCCCTGCTGCGACCG 516
Qy 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAAsPLeuThrIleAAsPArgSer 140
Db 517 TCAACCCCTGCTGGCCAGGGCCAGCGCTGGGAGTCTGAACCTCAACCAAGAACCGTAGC 576
Qy 141 LeuThrGlyLeuLeuGluProValSerProProSerProProHisThrAAsPProGluLeu 160
Db 577 CTGACGGGCAAGCTGGAACCGGTGTCTCCCGGAGCCCGGACACTGACCTGAGACTG 636
Qy 161 GluLeuValProProArgLeuSerIleGluGluLeuIleGlnAAsPProAAsPArgValAAsP 180
Db 637 GAGCTGGTCCGCGCAGCGCTCTCAAGAGAGACTGATCCCAACATGACCCGCTGAGAC 696
Qy 181 ArgGluIleThrMetValGluGlnGlnIleSerIleLeuLeuValIleGlnGlnGlnLeu 200
Db 697 CGAGAGATCACATGAGAGAGAGATCTTAAGCTGAAGAAGAGAGAGAGAGAGAGCTG 756
Qy 201 GluGluGluAlaAlaValProProGluProGluProValSerProProProIleGlu 220
Db 757 GAGAGAGAGGCTGCCAAGCGCGCCGAGCTGAGAAACCCGCTGACCGCGCCATCAGAG 816

QY 221 SerLeuHisArgSerLeuValGlnIleIleTyrAspGluAsnArgIleValaGluAla 240
DB 817 TCAGACACCGCAGCTGGTGAGATCATCTACGACGAGAACCGGAGAGGCTGAAGCT 876
QY 241 ALHHisArgIleLeuGluGluLeuGluProGlnValGluLeuProLeuTyrAsnGlnPro 260
DB 877 GCACATCGAATTTCTGGAAAGGCTTGAGGCCCCCAAGGTGAGCTGCCCTGTTCACACGACC 936
QY 261 SerAspThrArgGlnTyrHisGluAsnIleLeuIleAsnGlnIleMetArgIleValLeu 280
DB 937 TCCGACACCGGAGATCATGAGAACATCAAAATAAACGAGCGATCGCGAAGAAAGCTA 996
QY 281 IleLeuTyrPheLeuSarArgIleAsnHisAlaArgIleGlnIleProIleValPheCysGln 300
DB 997 ATCTTGATCTTCAAGAGGAGAAATCAGCTCGGAAACAAATGGAGCGAGAAATTTCTGTCAG 1056
QY 301 ArgTyrAspGlnLeuMetGluAlaLeuGluValLeuValGluValArgIleGluAsnAsnPro 320
DB 1057 CGCTATGACCAAGTCTCATGAGGCTTGAGAGAAAGGTGAGGCGCATCGAAACACCCC 1116
QY 321 ArgArgArgAlaLeuSerLeuValArgIleTyrTyrGluIleValPheProGluIle 340
DB 1117 CGCGGGGGGCCAAGAGAGCAAGGTTCCGAGTACTAGAGAAAGAGTTCCCTGAGATC 1176
QY 341 ArgIleGlnArgGluLeuGlnIleValArgMetGlnSerArgValGluIleValArgIleSerGly 360
DB 1177 CGGACGAGCGCAGCTGCGAGAGCGCATGCGAG---AGGGTGGGCCCGCGGCGAGTGGG 1233
QY 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
DB 1234 CTGTTCATGTCCGCCGCCGCGAGCGACGAGGTGTCAAGATCATCATGAGCTCTCA 1293
QY 381 GluGlnGluAsnLeuGluIleValArgIleMetArgIleValIleProPheMetLeuTyr 400
DB 1294 GACACGAGAGAACCTCGAGAGAGAGATGCGCAGCTGGCCGTGATCCGCCCATGCTGTAC 1353
QY 401 AsnAlaAspGlnGlnArgIleValPheIleAsnMetCysnGlyLeuMetAlaAspPheMet 420
DB 1354 GAGCGTGAACGAGCGCATCAAGTTCATCATGAAATGAAAGCGGCTTATGGCGGACCCCATG 1413
QY 421 LysValTyrIleAspArgGlnValIleMetAsnMetTyrSerGluGlnGlnIleValPhe 440
DB 1414 AAGGTGTAACAAGACGCCGAGTCAATGAAATGTGAGTGAAGAGAGAGAAAGAACTTTC 1473
QY 441 ArgGluIlePheMetGlnHisProIleValAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
DB 1474 CGGAGAGAGTTCAAGCATCCCAAGAACTTGGCCTGATCGCATCTTCGAGAGAG 1533
QY 461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrIleValAsnGluAsnTyrLys 480
DB 1534 AAGACAGTGGCTGAGTGGCTCTTATTACTTACTGACTTAAGAAATGAAATGAACTATAG 1593
QY 481 SerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGln 500
DB 1594 AGCTTGATGAGACGAGCTTACGGCCGCGGCAAGAGCCAGACGACAAACACGACAG 1653
QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
DB 1654 CACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1713
QY 521 AsnGluIleValGluIleValGluIleValGluIleValGluIleValGluIleValGlu 540
DB 1714 GAGGAG 1773
QY 541 AsnIleValArgIleValArgIleValArgIleValArgIleValArgIleValArgIle 560
DB 1774 GACAG 1833
QY 561 LysGluAlaValAlaSerIleValArgIleValArgIleValArgIleValArgIleVal 580
DB 1834 AAGAGAGCTGTGCTCTCAAGAGCGGCAAAATCGCAACAGCCAGGAGAAACCCAAAGC 1893
QY 581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluAlaIleThrProGlnGln 600

DB 1894 CGCATCACCGCTCAATGAGCTTAATGAGGCTCAACAGAGAGAGGCTCATCCCCGACAG 1953
QY 601 SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTyrThrGluGluMet 620
DB 1954 AGCGCCGAGCTGCTCCATGAGCTGAATGAGAGTTCTCGCTGACAGAGAGAGAGAG 2013
QY 621 GluThrAlaValArgIleValLeuGluHisGlyArgAsnTyrSerAlaIleAlaArgMet 640
DB 2014 GAAACAGCCAGAGAGAGTCTCTGAGAACCGCGCCGACATGCGTCCGCTCCGAGATG 2073
QY 641 ValGlySerIleThrAlaSerGlnCysIleValAsnPheTyrPheAsnTyrIleValArgGln 660
DB 2074 GTGGGCTCCAGACTGTGTGAGTGAAGAACTTACTTCACTTAAGAGAGAGAGAG 2133
QY 661 AsnLeuAspGluIleLeuGlnGlnHisIleValLeuValMetGluValArgAsnAlaArg 680
DB 2134 AACCTGATGAGATCTTCAAGACGACCAAGCTGAAGATGAGAGAGAGAGAGAGAGAG 2193
QY 681 ArgIleValValValAlaProAlaAlaAlaSerGluGluAlaAlaPheProValVal 700
DB 2194 AGGAG 2253
QY 701 GluAspGluGluMetGluAlaSerGlyValSerGlyValSerGluGluMetValGluGlu 720
DB 2254 GAGGATGAG 2313
QY 721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyValCysSerGlyProAla 740
DB 2314 GCTGAA-----GGC 2322
QY 741 ThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaVal 760
DB 2323 ACTGTCAACAAAGCTGACAGACCGGAGACATCCCTCTCTCACTAGAGCGGCAAG 2382
QY 761 AsnThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProPro 780
DB 2383 GACACAGGCGAGATGGGCCCAAGCCCGACGACCTTGAGGCGGAGAGAGAGAGAGAG 2442
QY 781 GlyProProThrProPro-ArgArgThrSerArgAlaProIleGluProThrProAlaSe 800
DB 2443 GGGCCACCAACCCCAACCAACGAGAGACATCCGGCC-CCCATGAGTCCACCCGGGCTC 2501
QY 800 rGluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProProPr 820
DB 2502 TGAAGCCACTTAAGCCCTTAAGCCCCCAAGAGACCCCATTTCCCTCTCACTCTCC 2561
QY 820 oValValProLysGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 840
DB 2562 TGTGTCTCCCAAG 2621
QY 840 yGluGluGluValProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluG 860
DB 2622 GAG 2681
QY 860 uProValLysSerGluCysThrGluGluAlaGluGluGluGluGluGluGluGluGlu 880
DB 2682 GCCCTCAAGAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2741
QY 880 aGluAlaAlaGluAlaThrAlaGluGluValAlaLeuValAsnIleValValGluGluG 900
DB 2742 GAGGCGCTGAGAGCCAG 2801
QY 900 rGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaTh 920
DB 2802 CGGAGAGGCGCACACAGCCAGAGAGCTCGGGCCGCCCGACGAGAGAGAGAGAGAGAG 2861
QY 920 rCysSerAlaAspGluValAspGluAlaGluGluGlyValAsnIleValAsnArgLeuLeuSerPr 940
DB 2862 CTGAGATGACAGAGAGTGAATGAAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2921
QY 940 oArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysPr 960

Db	2932	MAAGGCCAGCCTCTCCACCCCGACCTGGGAGACCCCGGGCCAAATGCTTCAACCCAGAAAGCC	2981
QY	960	OLEuAaPLeuLYSGInLeuLYSGInArgAAlaAlaIleProFolIeGInValThLY	980
Db	2982	ACTGGACCTGAAGCAGCTGAAGCAGCAGCGGCTGCATCCCGCCATCTCAGTCAACCA	3041
QY	980	gValHISGLuPProPAArGGLuAspAlaAlaProThrySPRoAlaPProPAArProP	1000
Db	3042	AGTCCTATAGCCCCCGGAGAGAGCAGCTCCCAACGACAGCTCCCGACGCCCAAC	3101
QY	1000	oProProGInaMeuLeuGInProGluSerAspAlaProGInGInPProGlySerSerProAr	1020
Db	3102	GCCACCGCAAAACCTGCAAGCGGAGAGGAGAGCCCTTACAGACCTGGCAGAGACGCCCG	3161
QY	1020	gGlyLYSerSerArgSerProAlaProProAlaAspLYSGlu-----	1033
Db	3162	GGGCAAGAGCAGAGACCCCGGACCCCGCGGACAGAGAGGACAGAGCCGTGTTCTT	3221
QY	1034	----AlaPheAlaAlaGluAlaGInLYLeuLeuProGlyLYAspProProCYSTPThrSerGI	1052
Db	3222	CCCAAGCCTTCGAGCGCAGCGGCCCAAGAACCTGCTGGAGACCCCTTCTGACCTTCGG	3281
QY	1052	LYLeuProPheProValPProPAArGGLuValILeLYaLYaSerProHISaIaPAArAspPr	1072
Db	3282	CCTGGCCCTTCCCCCGGCCCGCGAGGTATCAAGGCTCTCCCGCATGCCCCGAGACC	3341
QY	1072	oSerAlaPheSerTYraIaPProProGlyHISaProLeuProLeuGlyLeuHISaAPThraI	1092
Db	3342	CTCAGCCTTCTCTACGCTCCACCTGGTCAACCACTGCCCTTGAGCTCCATGACCTGC	3401
QY	1092	AArGProValLeuProArGProProThryIleSerAnbProBProLeuIleSerSerAl	1112
Db	3402	CCGGGCCCTCTCGCGCCGCCCAACCACTCATCTCCAAACCGCTCCCTCATCTCTCTGC	3461
QY	1112	ALyHISaProSerValLeuGluLYaGGLInIleGlyAlaIleSerGInGlyMeSerValGI	1132
Db	3462	CAAGACACCCCAAGCGCTCTCGAAGGCAAAATGAGTGCATCTCCCAAGGAATGTGGTCCA	3521
QY	1132	nLeuHISaValProTYrSerGInuHISaLYaLYaAlaProValGlyProValThrMetGlyLe	1152
Db	3522	GCTCCAGCTCCCGTACTAGAGCATGCCAAGGCCCGGTGGGCCCTGTCACTAGTGGGCT	3581
QY	1152	uProLeuBProMetArPProLYaLYaLeuAlaProPheSerGIYValLYSGInGluInLe	1172
Db	3582	GCCCTCGCCCATGAGCCCCCAAAAGCTGGACCTTCAGCGGAGTGAAGCAGCAGAGCT	3641
QY	1172	uSerProArGGLyGInaLYaGlyProProGluSerLeuGlyValaProThraIaGInGluAl	1192
Db	3642	GTCGCCAAGGGGGCCAGGCTGGGCGCACCCGAAAGGCTGGGGGTGGCCACAGCCCAAGAGGC	3701
QY	1192	aSerValLeuArGGLyThraIaLeuGlySerValProGlyLYaSerIleThryLYSGlyIL	1212
Db	3702	GTCGGTGTGAAGGAGACAGCTCTGGGCTCAGTTCCGGGCGGAAAGCATCACCAAAAGGCAT	3761
QY	1212	eProSerThraArgValProSerArPSerAlaIleThryTYrArgLYaSerIleThrHISGI	1232
Db	3762	TCCCAAGCACAAGGGTGGCCCTCGAAGAGGCATACATACCGGGCTTCATACACCAAGG	3821
QY	1232	TYThProAlaAspValLeuTYrLYaGlyThryIleThraArgIleIleGlyIuAspSerPr	1252
Db	3822	CACCCCAAGCTGACCTCTGCTCAAGAGGCACCATCACAGAGTCACTGGCGAGGACACCC	3881
QY	1252	oSerArgLeuAspArGGLyArGGLuAspSerLeuProLYaLYaHISaValIleTYrGluG	1272
Db	3882	GAGTCGCTTGACCGCGGCGGAGAGCAGCTTCCCAAGGGCCACGTCACTACGAAAGG	3941
QY	1272	LYLYaLYaGlyHISaValLeuSerTYrGluGlyLYaMeSerValThrGInCYaSerLYSGI	1292
Db	3942	CAABAAGGCCACAGCTTGTCTCTATGAAGGTGGCATGTCTGTGACCAAGTCTCCAAAGCA	4001
QY	1292	LYAspGlyLYaArgSerSerGlyProProHISaGluThraAlaProLYaArGThryTYraS	1312
Db	4002	GGAGCGCGAAGCAGCTCAGAGACCCCGCCATGACAGCGGCGCGCCCAAGCGCACCTATGA	4061

OY	1312	pMecMetCgluGlyArgValGlyArgAlaIleSerSerIleSerIleGluGlyLeuMetCgl	1332
Db	4062	CATGATGAGAGGGCCCGCTGGGGAGAGGCATCTCTCCACCGAGCATCGAAGGCTCATGGG	4122
OY	1332	YArgAlaIleProProGluArgGHisSerProHisIleLeuGlyGluGluHisIleIleAr	1352
Db	4122	CCGTGCCATCCGGCCGGAGCGACACAGCCCCCACCACTTCAGAGGACACACATCCG	4182
OY	1352	GglySerIleThrGlnGlyIleProArgSerTyValGluIleGlnGluIleuArgTyLeuAr	1372
Db	4182	CGGGCTCATCACACAAAGGATCCCTCGGTCTCACTGGAGGCACAGAGGACTACTCCG	4242
OY	1372	GArgGluAlaTybLeuLeuGlybArgGluGlyTyThrProProProProProSerArgAs	1392
Db	4242	TGGAGAGGCCAAGCTCTTAAAGCCGGAGGAGGAGCCCTCCGGCCCCACCGCTTCACGGGA	4302
OY	1392	PLeuThrGlnuIleATyIleThrGlnAlaLeuGlyProLeuTybLeuTybProAlaHisGly	1412
Db	4302	CCTGACCGAAGGCTCAAGAGGAGCGCCCTGGCCCCCTGAAGCTGAAGCCGGCCATCA	4362
OY	1412	uGlyLeuValAlaThrValTybGluAlaGlyArgSerIleHisGluIleProArgGluGly	1432
Db	4362	GGGCTGGTGGCCACCGGTGAAGAGAGCGCGCGCTCCATCATGATAGATCCCGCCGAGGA	4422
OY	1432	uLeuArgHisThrProGluuLeuProLeuAlaProArgProLeuTybGluGlySerIleTh	1452
Db	4422	GCTGGGACACGCCCCGAGCTGCCCTTGAGCCCCGGCGGCTTCAGAGAGGCTCATAC	4482
OY	1452	rGlnGlyThrProLeuTyTyraaThrGlyAlaSerThrThxGlySerTybHisAs	1472
Db	4482	GCAAGGACCCCGCTCAAGTACGACACCGGCGCTCCACCTGGCTCCAAAAGACAGA	4542
OY	1472	pValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuArgVa	1492
Db	4542	CGTACGCTCTTCATCGGACAGCCCCGGCGGAGCTTCCACCGGTGACCCGCTGAGTGT	4602
OY	1492	IMeAlaAspAlaArgAlaLeuGluAlaArgAlaCyTyGlyuGlySerLeuTybSerArgPr	1512
Db	4602	GATGGCGAGCCCGGGGACCTGGAACGTGCTCTTACGAGGAGAGCTGAAGAGACCCGGCC	4662
OY	1512	oGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGly	1532
Db	4662	AGGACCGCCACGAGCTCGGGGGGCTCATTTGCCGGCGGCGCCCGGTCAATTGGCTGA	4722
OY	1532	uLeuGlyArgProArgInserProLeuThrTyGluAspHisGlyAlaProPheAlaGly	1552
Db	4722	GCTGGGTAAAGCCGGCGAGAGCCCCCTGACCTTAAGAGACACCGGGGACCCCTTGGCCGG	4782
OY	1552	YHisLeuProArgGlySerProValIleMetArgGluProThrProArgLeuGlnGluGly	1572
Db	4782	CCACTCCACAGAGGTTCCGCCGAGACATGCGGGAGCCCAACCGCGCGCTCGACGAGGG	4842
OY	1572	ySerLeuSerSerSerTybAlaSerGlnAspArgTybLeuTherThrProArgGlyuIle	1592
Db	4842	CAGCCTTTTCGTCAGCAAGGATCTCCAGACCGAAGACTAGCTGACGCTCTTGAGAT	4902
OY	1592	eAlaTybSerProHisSerThrValProGluHisIleProHisIleProIleSerProTyGly	1612
Db	4902	CGCCAACTCCCGGACACACACCGTGGCCGAGACCAACCAACCCATCTCGCCCTTGA	4962
OY	1612	uHisLeuLeuArgGlyAlaSerGlyValaAspLeuTyArgSerHisIleProLeuAlaPh	1632
Db	4962	GCACCTGCTTCGGGGCTGAGTGGCGTGAACCTGTATCGAGCCACATCCCGCTGGGCTT	5022
OY	1632	eAspProThSerIleProArgGlyuIleProLeuAspAlaAlaAlaIleATyTyTyLeuAspR	1652
Db	5022	CGACCCCACTCCATACCCCGCGGACTCCCTCTGAGACGACCGCTGCTCACTACCTGCC	5082
OY	1652	oArgHisLeuAlaProAsnProThrTyProHisIleuTyTyProProTyTyLeuIleArgIle	1672
Db	5082	CCGACACCTGGCCCCCAACCCCACTACCCGACCTGTACCAACCTTACCTCATCTCCGGG	5142

QY 1672 yTyPrOaBpThrAlaAlaLeuGluaBnaRgInThrIleIaAsnaPtyrIleThrSe 1692
Db 5142 CTACCCCGACACGGCGCGCTGGAGAAACGGCAGACCATCATCATGACTACATCACTC 5201
QY 1692 rGIngluMetHisHisAsnThrAlaThrAlaMetAlaGlnARgAlaAspMetLeuArgI 1712
Db 5202 GCGAGAGATGCACACAAACACGGCCACCGCATGGCCACGAGCTGATATGCTGAGGG 5261
QY 1712 yLeuSerProARgIuSerSerLeuAlaLeuAenTYrAlaAlaGlyProARgIyIleI 1732
Db 5262 CCTCTGCCCCGGAGATCTCGCTGGACATCACTACGCTGGGGGTCCCGAGGCACTCAT 5321
QY 1732 eAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAl 1752
Db 5322 CGACCTGTCCCAAGTCCACACCTGCTGTGCTCGTGGCCCCGACACAGGACCCCAAGC 5381
QY 1752 aThrAlaMetAspARgLeuAlaTYrLeuProThrAlaProGlnProPheserArghI 1772
Db 5382 CACCGCCATGAGACCGCTTGCTTACCTCCACCGCGCCCAAGCCCTTCAGCAGCCGCA 5441
QY 1772 sSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrSe 1792
Db 5442 CAGCACTCCCACTCTCCCAAGAGGTCCACACACTTGACAAACCAACCAACCACTC 5501
QY 1792 rSerSerGluARgGluARgAspARgAspARgIuARgAspARgIuARgIuIy 1812
Db 5502 CTCTGTCGAGCGGAGCAGAGCCGGATCGAGACCGGAGATCGGGATCGGGAGCGGAAAA 5561
QY 1812 sSerIleuThrSerThrThrThrValGlnHisAlaProIleThrARgProGlyThrG 1832
Db 5562 GTCAATCTCATTACATCCACACAGCGGAGAGCAGCACTCATGAGACTGATACAG 5621
QY 1832 uGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerArProAl 1852
Db 5622 GCGAGACAGGGGAGCAGCAGCGCAGCAGCGCGGGGGGGGGGAGCAGCAGCGCCCGC 5681
QY 1852 aSerHisSerHisAlaHisGlnHisSerProIleSerProARgThrGlnAspAlaLeuG 1872
Db 5682 CTCCCACTCCCATGCCACACAGCACTCGCCATCTCCCTCGGACCCAGGATCCCTCCA 5741
QY 1872 ngInaRgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluPr 1892
Db 5742 GCGAGAACCCAGTGTCTTACAAACAGGCAATGAAGGATATCATCCGCTTGAGGCC 5801
QY 1892 oSerLysProThrValLeuARgSerThrSerThrSerSerProValARgProAlaIaTh 1912
Db 5802 CAGCAAGCCACGAGTCTGTAGGTCACCTCACTCTCACTCCGCTTCGCCAAGCTGCCAC 5861
QY 1912 rPheProProAlaThrHisCysARgLeuGlyGlyThrLeuAspGlyValTyPrProThrIe 1932
Db 5862 ATTCACCACTGCACCACTGCTGAGGCGGACCTCGATGGGGTCTAACCTTACCT 5921
QY 1932 uMetGluProValLeuLeuProIySGluAlaProARgValAlaARgProGluARgProAr 1952
Db 5922 CATGGAGCCGCTTGTCTGTCCAAAGAGGCGCCCGGGGTGCCCGGCAAGGCGCCCG 5981
QY 1952 gaIaAspThrGlyHisAlaPheLeuAlaLysProProAlaARgSerGlyLeuGluProAl 1972
Db 5982 AGCAGACACCGGCAATGCTTCTTCGCCAAGCCCAAGCCCGCTCCGGGTGAGCCCGC 6041
QY 1972 aSerSerProSerLysGlySerGluProARgProLeuValProProValSerGlyHisAl 1992
Db 6042 CTCTCTCCCCCAGAGAGGCTCGGAGGCCCGGCCCCCTAGTGCCTCTCTGGCCACGC 6101
QY 1992 aThrIleAlaARgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProPr 2012
Db 6102 CACCATGCGCGCACCCCTGCGAAGAACTTCGACCTCACACAGCCAGCCGGAACCGGCC 6161
QY 2012 oAlaProAlaSerAlaSerAspProHisARgGlyLysThrGlnSerLysARgProPhe 2032
Db 6162 GGGGCCACCTGCTCGGCTCGGACCGGACCGGAGAAAGACTCAAAAGTAAACCTTTTC 6221
QY 2032 rIleGlnGluLeuGluLeuARgSerLeuGlyTyHisGlySerSerTySerProGluG 2052

Db 6222 CATCCAGAACTGGAACTCCGTTCTTGGGTTACACCGCAGACAGTATACGCCCGAAAG 6281
QY 2052 yValGluProValIserProValIserSerProSerLeuThrHisAspLysGlyLeuProLy 2072
Db 6282 GGTGAGGCCCTTCAGCCCTGTGACTCACCCAGCTGACCCACGACCAAGAGGGCTCCCA 6341
QY 2072 sHisLeuGluGluLeuAspLysSerHisLeuGluGlyLeuLeuARgProLyGlnProG 2092
Db 6342 GCACTGGAAGAGCTGCACAAAGGCCACTGAGAGGGGAGCTGCGGCCCAACAGCAGCAAG 6401
QY 2092 yProValLysLeuGlyGlyGluAlaHisLeuProHisLeuARgProLeuProGluSe 2112
Db 6402 CCGCGTGAAGCTTGGGAGAGCGCGCCCACTCCACACTGGCGCCGCTGAGAG 6461
QY 2112 rGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnARgVa 2132
Db 6462 CAGCCCTGTCTCAGCCCGCTGCTCCAGACCCGCCAGGGGTCAAAAGTCAACAGCGG 6521
QY 2132 lValThrLeuAlaGlnHisIleSerGlyValIleThrGlnAspTYrThrARgHisHisPr 2152
Db 6522 GGTCAACCTTGCGCCAGCATATCATGAGGTATCAACAGACTACACCCGACACACCC 6581
QY 2152 oGlnGlnLeuSerAlaProLeuProAlaProLeuTYrSerPheProGlyValAserCysPr 2172
Db 6582 ACAGAGCTTCAGCGCACCTCTCGCGCCCTCTTACTCTTCTCGGGGCCAGCTGCC 6641
QY 2172 oValLeuAspLeuARgARgProProSerAspLeuTYrLeuProProProAspHisGlyAl 2192
Db 6642 CGTCTGAGACTCCCGCGCCCAACCAAGTACCTTACCTCCGCGCCCGGACCATGTGTC 6701
QY 2192 aProAlaARgGlySerProHisSerGlyGlyLysARgSerProGluProAsnLysTh 2212
Db 6702 CCGGGCCGTGCTCCCGCCCAAGCGAAGGGGCAAGAGTCTCCAGACCAAAACAAGAC 6761
QY 2212 rSerValLeuGlyGlyGlyGluAspGlyIleGluProValIserProProGluGluMetTh 2232
Db 6762 GTCTGTGTTGGTGGTGTGAGAGACGATTAATCACTGTCTCCACCGAGGAGCATGAC 6821
QY 2232 rGluProGlyHisIserARgSerAlaValTYrProLeuLeuTYrARgAspGlyGluGlnTh 2252
Db 6822 GAGGCCAGGGCACTCCCGAGTGTGTGTACCCGCTGTCACCGGAGGGGAACAGAC 6881
QY 2252 rGluProSerARgMetGlySerLysSerProGlyValanThrSerGlnProProAlaPhePh 2272
Db 6882 GGAGCCACAGAGATGGCTCCAAAGTCTCAGGCAACACACAGCCCGCAGCTTCTT 6941
QY 2272 eSerLysLeuThrGlnuSerAsnSerAlaMetValLysSerLysGlnGluIleAsnLy 2292
Db 6942 CAGCAAGCTGACCGAGAGCACTCGCCATGTGTCAAGTCAAGAAAGCAAGAGATCAACA 7001
QY 2292 sLysLeuAenThrHisAsnARgAsnGluProGluTYrAenIleSerGlnProGlyThrG 2312
Db 7002 GAAAGTGAACACCCACAAACCGAATGAGCTGAATACAGTACAGCAGCTGAGCGGA 7061
QY 2312 uIlePheAenMetProAlaIleThrGlyThrGlyLeuMetThTYrARgSerGlnAlaVa 2332
Db 7062 GATCTTCAATATGCGCGCATACCGGAACAGGCTTATGACTATGAAGCAAGCGCGT 7121
QY 2332 lGlnGlnHisAlaSerThrAsnMetGlyLeuGlnAlaIleIleAsnGlyValaLeuMetG 2352
Db 7122 GAGAGAACATGCGCAGACCAACATGGGGCTGAGGCGCAATTTAGAAAGCACTAGGG 7181
QY 2352 yLysTYrAspGlnTYrGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAs 2372
Db 7182 TAAATATGACAGTGGGAAGAGTCCCCGCGCTCAGCGCCAAATGCTTTAACTCTGAA 7241
QY 2372 nAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyARgSerAspHis 2392
Db 7242 TGCAGATGCGACGTCGCGGTATGCTATGCCAATACCGCTGTGACGAGCAGGAGTGA 7301
QY 2392 sThrLeuThrSerProGlyGlyGlyValaLysValIserGlyARgProSerSerAr 2412

Db 7302 CACACTACCTGCGCAGGTGGCGGGAGGCCAAGGCTTCTGCGAGACCACGACCG 7361
Qy 2412 GLYEALALYSERPROALAPROGLYLEUALASERGLYASPARPROPOSERVALSERSE 2432
Db 7362 AAAAGCCAAAGTCCCCGGGCGGCTGGCATCTGGGAGACGGGCACTCTGCTCTTC 7421
Qy 2432 TVALHISERGLUGLYAPRYASAMARGARGTHRPROLEUTHRANARGVALTPGLUAS 2452
Db 7422 AGTGCACTCGGAGGAGACTGCAACCGCGGACCGCTCAACACCGCTGGGAGGA 7481
Qy 2452 PARGPROSERSEALAGLYSERTHRRPHROPTRYRANPROLEULEMETAGLEUGL 2472
Db 7482 CAGGCGCTCGTCCGAGGTCCAGCATTCCTTACAACTCCGTATCATGCGGTGCA 7541
Qy 2472 PALAGLYVALMETALASERPROPROPROGLYLEUPROALAGLYSERGLYPROLEUA 2492
Db 7542 GCGGCGTCTCATGGCTTCCCAACCCCGGCGCTCCCGGCGGACGGGCGCTTCGC 7601
Qy 2492 AGLYPROHISERHISALATRAPARGUGLUPROLYAPROLEULEUCYSSERGLYRGJUTH 2512
Db 7602 TGGCCCCCACCACCGCTGGAGAGAGGCCAACGCACTGCTGCTGCGAGTACGAGAC 7661
Qy 2512 RLEUSERAPSERGLU 2517
Db 7662 ACTCTCCGACAGCGAG 7677

RESULT 11

ADL13811
ID ADL13811 standard; DNA, 8686 BP.

AC ADL13811;

XX 06-MAY-2004 (first entry)

DE Osteoarthritis-associated polymorphic nucleotide #343.

XX de; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;

KM joint space narrowing; osteophyte development; joint pain;

XX osteoarthritis; SNP; single nucleotide polymorphism.

OS Homo sapiens.

XX MO2003054166-A2.

XX 03-JUL-2003.

PF 19-DEC-2002; 2002MO-US041225.

XX 20-DEC-2001; 2001US-0342603P.

XX (INCY-) INCYTE GENOMICS INC.

XX Jones KA, Schaefer A;

XX WPI; 2003-559141/52.

XX Determining susceptibility of an individual to joint space narrowing,

XX osteophyte development and/or joint pain comprises identifying whether

XX the individual has at least one polymorphism in a polymorphonucleotide encoding

XX a protein.

XX Disclosure; SEQ ID NO 343; 297bp; English.

XX The invention relates to a method of determining susceptibility of an

XX individual to joint space narrowing and/or osteophyte development and/or

XX joint pain comprising identifying whether the individual has at least one

XX polymorphism in a polymorphonucleotide encoding at least one of the protein

XX listed in the specification. The method, composition and agent are

XX useful for modulating the susceptibility of an individual to joint space

XX narrowing and/or osteophyte development and/or joint pain that is

XX associated with a disease, preferably osteoarthritis. The cell line and

XX the non-human animal are useful for screening for an agent for diagnosing

XX an individual having susceptibility to joint space narrowing and/or

CC osteophyte development and/or joint pain. This sequence corresponds to
CC the polymorphonucleotide encoding a protein listed in the specification. (Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pat_sequences).

XX SO Sequence 8686 BP; 1881 A; 3069 C; 2562 G; 1174 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0	Length:	8686
Score:	12978.00	Matches:	2487
Percent Similarity:	98.57%	Conservative:	3
Best Local Similarity:	98.46%	Mismatches:	9
Query Match:	98.21%	Indels:	28
DB:	10	Gaps:	3

US-09-522-753-5 (1-2517) x ADL13811 (1-8686)

Qy 1 MetSerGlySerThrglnLeuValAlaGlnThrTrpArgAlaThrglnProArgTyrPro 20
Db 157 ATGTGGGCTCCACAGAGCTGTGGACAGAGCGGAGGCGCACTGAGCCCGCTACCGG 216
Qy 21 ProHISerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
Db 217 CCCACAGCCCTTCTTACCCAGTGCAGATGCGCCGAGCAGACAGCGAGCTGGGCTCTG 276
Qy 41 GluTyrGlnHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
Db 277 GAGTACGAGACCACTCCGAGCTATGCTCCACCTGTGGCCGCTCCATCATCAG 336
Qy 61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyIleAngIleArgSerGln 80
Db 337 CCCAGGCGGAGAGCGCTCTCTGCTGTGAGTTCAGCGCCGAGATGAAGCTCCAG 396
Qy 81 GluLeuHisLeuArgProGlnSerHisSerTyrLeuProGlnLeuGlyIleSerGluMet 100
Db 397 GAGCTCCACCTGGCGGCGCAGAGTCCACTCATCTGCCCCAGCTGGGAACTCAGATG 456
Qy 101 GluPheIleGlnSerIleArgProArgGlnLeuLeuProAspProLeuLeuArgPro 120
Db 457 GAGTCAATTAAGCAAGCGCCCTCGCTAGAGCTGTGCTGACCCCTGCTGCGACG 516
Qy 121 SerProLeuLeuAlaThrglnProAlaGlySerGlnAspLeuThrIleAspArgSer 140
Db 517 TCAACCCCTGTGCGACAGGCGCAGCTGCGGATCTGAACCTCAACCAAGACCGTAGC 576
Qy 141 LeuThrGlyIleLeuGlnProValSerProProSerProProHisThrAspProGlnLeu 160
Db 577 CTGACGGGCAAGCTGGAAACGGGTGTCTCCCGCAGCCCGCCGACACTGAGCTG 636
Qy 161 GluLeuValProProArgLeuSerIleGlnIleGlnIleGlnMetAspArgValAsp 180
Db 637 GAGCTGGTGGCGGCGCTGTCCAGAGAACTGATCCAGAACTGAGCCCGGTGAC 696
Qy 181 ArgGlnIleThrMetValGlnGlnGlnIleSerIleLeuIleValIleValGlnIleLeu 200
Db 697 CGAGAGATCCACTGTGTAAGCAGCAGATCTTAAGTGAAGAAGCAGCAAGCTG 756
Qy 201 GluGlnGlnAlaAlaIleProProGlnProGlnIleProValSerProProIleGln 220
Db 757 GAGGAGGAGGCTGCCAAGCGCGCCGAGCTGAGAAACCCCTGTACCGCGCCCATCGAG 816
Qy 221 SerIleHisArgSerLeuValGlnIleIleTyrAspGlnAspArgIleValIleGlnAla 240
Db 817 TCGAAGACCGGACGCTGGTGCAATCATCTACGAGAGAACCGAAGAGGCTGAAGCT 876
Qy 241 AlaHisArgIleLeuGlnGlyLeuGlyProGlnIleValGlnLeuProLeuTyrIleAngIlePro 260
Db 877 GCAATCGAATTTGGAAGGCTGGGCGCCAGAGTGAAGTGCCTGTATCAACAGCC 936
Qy 261 SerAspThrArgGlnTyrHisGlnIleValIleValIleGlnIleAlaMetArgIleValLeu 280
Db 937 TCCGACACCGCGGAGTATCATGAGAACTCAATAATTAACAGCGCATGCGGAGAAAGCTA 996

OY	281	11LeuTyrPheIyEaYgArGAsnHiEa1AaArgLySgIntPlybGlnLyPheCyGln	300
Db	997	ATCTTGAATCTCAAGAGAGAAATCAACCTTCGGAAACAATGGAGCAAAATTCTGCAG	1056
OY	301	ArgTyrAspGlnLeuMetGluAlaLeuGlnLybSybVa1GluArg1LeG1AaAsnPro	320
Db	1057	CGCTATGACCAAGCTCTATGGAGGCTTCGGAGAAAGAGTGGAGCCCTCGAACAACCC	1116
OY	321	ArgAaArgAlaLybGlybSerLybVa1ArgGluTyrTyrgLybGlnPheProGluLe	340
Db	1117	CGGGGGCGGGCCAAAGAGAGCAAGGTTCCGACAGTCACTACGAAAGAGTTCCTCGAGATC	1176
OY	341	ArgLybGlnAcyGlnLeuGlnGlnAaPheMetGlnSerAaVa1Gly1AaArg1SerGly	360
Db	1177	CGCAAGCAAGCCGAGACTTCAGAGAGCGCATGCAAG--AGAGTGGCCAGCGGAGCATGGG	1233
OY	361	LeuSerMetSerAlaAlaArgSerGlnHiGlybVa1SerGluLe1eAspGlyLeuSer	380
Db	1234	CTGTTCATGTCGCGCCGCCGACGAGACAGAGGTGTACAGATCATGATGCTGCTCA	1293
OY	381	GluGlnGlnuAaenLeuGlnLybGlnMetAaArgGlnLeuAlaVal11eProProMetLeuTyr	400
Db	1294	GAGCAGGAGAACTTGGAGAAACAGATGGCCAGCTGGCTGGATCCGCCATCTGTAC	1353
OY	401	AspAlaAspGlnGlnAaArg1LeLybPhe1LeaMetAaenGlyLeuMetAlaAspProMet	420
Db	1354	GACCGTGACCAAGCAGCGCATCAAGTTCATCAACATGAACGGGCTTATGGCCGACCCATG	1413
OY	421	LybVa1TyrybAspAaArg1Va1MetAaMetTPSerGlnGlnLybGlnTybPhe	440
Db	1414	AAGGTGTCAAAAGCCGCCAGGTCTATGAATGTGGAGTGACAGAGAAAGAACCTTC	1473
OY	441	ArgGlnLybPheMetGlnHiProLybAsnPheGlyLeu11eAaSerPheLeuGluArg	460
Db	1474	CGGAGAAATTCAATGACACATCCCAAGAACTTTGGCTTGATGCAATCTCTGGAGAGG	1533
OY	461	LybThrVa1AlaGlnCybVa1LeuTyrTyrybLeuThrLybLybAsnGluAaSnTyrys	480
Db	1534	AAGACAGGCTGAGTGCTCTCTATTACTACGACTTAAGAAAGAAATGAACATATAG	1593
OY	481	SerLeuVa1AaArgAaSerTybArgAaArgGlybSerGlnGlnGlnGlnGln	500
Db	1594	AGCTGTGTAACCGAGCTATTCGGCGCGGCAAGAACCCAGCAACAACAGCAGAG	1653
OY	501	GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProAaSerSerGlnGlnGlnLybS	520
Db	1654	CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCCATGCCCGCAGAGCCAGAGAGAGAA	1713
OY	521	AspGlnLybGlnLybGlnLybGlnAlaGluLybGlnGlnGlnLybProGluVa1GluAa	540
Db	1714	GATGAGAAAGAGAAAGAAAGAGCGGAGAGAGAGAGAAAGCCGAGGTGGAGAAC	1773
OY	541	AspLybGlnAaPheLeuLeuLybGlnLybThrAspAaPThrSerGlyGluAaAsnAspGlu	560
Db	1774	GACAAAGAAAGCTCTCTCAAGAGAAAGCAAGCAGCACTTAGGGAGAGACAAAGCAGAG	1833
OY	561	LybGlnAlaVa1AaSerLybGlybArgLybThrAlaAaenSerGlnLybArgLybGly	580
Db	1834	AAGAGAGCTGTGGCTCCAAAGAGCGCAAACTGCCCAACAGCCAGGAGAAAGCAAAAGCC	1893
OY	581	ArgGlnThrAaArgSerMetAlaAsnGlnAlaAaenSerGlnGluAla11eThrProGlnGln	600
Db	1894	CGCATCACCCCTCATATGATAGGCAACAGCGAGAGAGCCATACCCCCACAGAG	1953
OY	601	SerAlaGlnLeuAlaSerMetGlnLeuAaenGlnSerSerAaArgTPThrGlnGlnGlnMet	620
Db	1954	AGCGCCGAGCTGGCTTCATGAGAGCTGATAGAGTTCTCGCTGACAGAAAGAAATAG	2013
OY	621	GluThrAlaLybLybGlyLeuLeuGlnuHiGlybArgAaenTPSerAla11eAlaArgMet	640
Db	2014	GAACACGCAAGAAAGCTCTCTCGAACAAGCGCCGCAACTGATGGCCATGCGCCGATG	2073

OY	641	VAlGlySerIySthrValISerGIcYelYAsmPheryrPheAenTyrlYlYvAAgIn	660
Db	2074	GtGGGCTCCCAAGACGTGTGTGGCAGTGAAGAATCTTAACCTCAACTCAAGAAAGGCAG	2133
OY	661	AAsnLeuASpGluITeLeuGIInGlnHILyISylLeuYSmetGluYSGLIArGASpAlaArG	680
Db	2134	AACTTCATGATGAACTTCTCGACGACACAAGCTGAAGATTGAGAAAGAGAAAGCACGCCGG	2193
OY	681	ArGlYsLySyLysLysAlaProAlaAlaASerGIuAlaAlaPheProProValI	700
Db	2194	AGGAAGAAAGAAAGAGCGCGCGCGCGCCAGCAGAGAGGCTGCATTCCCGCCGTGGTG	2253
OY	701	GIuASpGIuGIuMeTGluAlaASerGIYalSerGIYasngIGluGIuMeValIGluGIu	720
Db	2254	GAGATGTGGAGATGGAGGCGTCCGGCGGTGACGGAAATGAGGAGATGTGTAGAGAG	2313
OY	721	AlaGIuAlaLeuHIlaISerGIYAsnGIuValProArGIyGIuCySerGIYProAla	740
Db	2314	GCTGAH-----GCC	2322
OY	741	ThrValASnaASnSerASpThrGIuSerIlleProSerProHISThrGIuAlaAlaYs	760
Db	2323	ACTGTCAAACAACAGCTCAGACACCGAGAGCATCCCTCTCTCTCACACTGAGAGCGCCAAAG	2382
OY	761	ASpThrGIYInASngIYProLYsProProAlaThrlEugIYalaaSpGIYProProPro	780
Db	2383	GACACAGGGCGAAATGGGCCCCCACCACCCCTGGCGCGCAGAGGGCCACCCCACA	2442
OY	781	GIYProProThrProPro-ArgaagThrSerArgAlaProIIeGIuProThrProAlase	800
Db	2443	GGGCGACCCACCCACACCGAGAGACATCCGGCC-CCACTGAATCCACCCGGGCTCT	2501
OY	800	rGIuAlaThrGIYalAProThrProProProAlaProProSerProSerAlaProProPr	820
Db	2502	TGAAGCACCTTAAGCCCTTAGCGCCCCACACAGACCCCATTTCCCTTCACTTCCTCC	2561
OY	820	OvalValProLYSGluGIuLYSGluGIuThrAlaAlaLaProProValIGluGIuGI	840
Db	2562	TGTGGTCCCCAAGAGAGGAAGAGAGAGAACCCGACAGCGCCCCACATGGAGAGAGGG	2621
OY	840	YGIuGIuGIuLYSProProAlaalaglGIuLUeuaIVaIASpThrGIYLSalaglGIu	860
Db	2622	GGAGAGAGAGAAAGCCCCCGCGCTGAGAGACTGCACTGACACAGAGAAAGCCGAGAGA	2681
OY	860	uProValLYSserGIuCYThrGIuGIuAlaglGIuGIYProAlaLYSGIYLSASPAl	880
Db	2682	GCCCTCAAGAGCCAAGTGCACGAGAGAAAGCCGAGAGGGCGCCGCAAGGCCAAGAGCG	2741
OY	880	aGIuAlaAlaGIuAlaThralaglGIYalaleuLYsaIaGIuLYSYSGIuGIYSe	900
Db	2742	GGAAGCGCGCTGAGGCGCACGGCCGAGAGGGCGCTCAAGGCACAGAAAGAGGGCGGGAG	2801
OY	900	rGIYArGIAlaThrThralalySerSerGIYAlaProGIuASpSerASpSerAlaATH	920
Db	2802	CGGCAGGGGCCACACAGCCAAAGACTCGGGGGCCCCCAGGACAGCACTCAATGTCAC	2861
OY	920	rCYSerSeraIaSPGIuValaSPGIuAlaglGIYGIYASpLYSaanaRgleuLeuSERPr	940
Db	2862	CTGCAGTGCACACAGATGGATGTAGAGCCCGAGAGGGCGGGCAGCAAGAACCGCTGTCTCCC	2921
OY	940	oArGPProSerLeuLeuthrProThrGIYASpProArGIAlaAnaIaSerProGIuLYSPr	960
Db	2922	AAAGGCCACGCTCTCAACCCGACCTGGGAGACCCCGGGCCAAATGCTTCACCCAGAAAGCC	2981
OY	960	oleuASpLeuLYSGluLeuLYSGluARgalAlaAlaalleProProIIeginValThrlY	980
Db	2982	ACTGGAACCTGAAGAGCTGAAGAGCAGAGCGGCTGCATCCCGCCCACTCCAGATCAACAA	3041
OY	980	eValHIeGIuProProArGIuASpAlaAlaProThrLYSProAlaProProAlaProPr	1000
Db	3042	AGTCATATAGCCCCCGGAGGAGCGCAGCTCCACCAAGCACAGCTCCCGAGGCCACAC	3101
OY	1000	oProProGIuAnaLeuGIuInProGluSerASpAlaProGIuInleProGIYSerSerProAr	1020

Db 3102 GCCACCCCAAACTGACGCGGAGACGAGCCCTCAGACGCTGACAGACCCCG 3161
Qy 1020 GGLYLSerArGerProAlaProProAlaAspLyGlu----- 1033
Db 3162 GGGCAAGACAGAGAGCCCGGACCCCGGACAGAGAGGACAGAAAGCTGTCTT 3221
Qy 1034 ----AlaPheAlaAlaGluAlaGluLyLeuProGlyAspProProCysTrpThrSerG 1052
Db 3222 CCCAGCCTTGACAGCCGAGGCGCAGAAAGCTGCTGGGGAACCCCTTGCTGACCTTCCGG 3281
Qy 1052 YLeuProPheProValProProAlaGluValAlaLyAlaSerProAlaProAspPr 1072
Db 3282 CCGGCCCTTCCCGCTGCGCCCGCTGAGTGATCAAGGCTCCCGGATGCGCGAGCC 3341
Qy 1072 oSerAlaPheSerTrpAlaProProGlyAlaProLeuProLeuGlyLeuAlaAspThrAl 1092
Db 3342 CTCAGCCTTCTCTTAAGCTCACTGACCTGGTCAACCCTGCTGGGCTCCATGACCTGC 3401
Qy 1092 aATGProValLeuProAlaProProThrTrlSerAsnProProProLeuLleSerSerAl 1112
Db 3402 CCGGCGCGCTGCTGCGCGCCACCCACATCTCCAAACCGGCTCCCTCATCTCTTGC 3461
Qy 1112 aLYeHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValG 1132
Db 3462 CAAGCACCCAGAGCTCTCCAGAGGCAAAATAGTGCCATCTCCAGAAATGTCGTCCA 3521
Qy 1132 nLeuHisValProTrpSerGlnHisAlaLyValaProValGlyProValThrMetGlyLe 1152
Db 3522 GCTTCAGTCTCCGACTCAAGTCAAGTCCAAAGCCCGGAGGCTCTTCAATGAGGAGCT 3581
Qy 1152 uProLeuProMetAspProLySlyLeuAlaProPheSerGlyValLySglnGlnIle 1172
Db 3582 GCCCGTGCATGGAACCCAAAAGCTGGCACCTTCAAGGAGTGAAGACAGAGAGCT 3641
Qy 1172 uSerProArgGlyGlnAlaGlyProProGlySerLeuGlyValProThrAlaGlnGluAl 1192
Db 3642 GTCGCCACGGGGGCGAGGCTGGGCGCACCGAGAGGCTGGGGTGCCCAACCCAGAGAGC 3701
Qy 1192 aSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLySgIle 1212
Db 3702 GTCCGAGCTAGAGGACAGCTCTGGGCTCAATTCGAGGAGAACATCCAAAGGCAT 3761
Qy 1212 eProSerThrArgValProSerAspSerAlaIleThrArgGlySerIleThrHisG 1232
Db 3762 TCCAGACACGGGCTGCTCTGACAGCGGCATCACTACCGGCTCCATCACCCAGC 3821
Qy 1232 YThrProAlaAspValLeuTrpLySgIleThrIleThrArgIleIleGlyLysAspSerPr 1252
Db 3822 CACGCCAGCTGACGTCTGTAAAGGGCACATCAACAGGATCATCGCGAGAGACAGCC 3881
Qy 1252 oSerArgLeuAspArgGlyArgGlyuAspSerLeuProLySgIleHisAlaIleTrpGlu 1272
Db 3882 GAATCGCTTAGACCGGCGCGGAGACAGCTGCCCCAAGGGCACCTCACTCAAGAGG 3941
Qy 1272 YLySlySgIlyHisValLeuSerTrpGluGlyGlyMetSerValThrGlnCysSerLySg 1292
Db 3942 CAAGAAAGGCGCAGTCTGTCTATAGGGTGGCATCTCTGACCCAGGCTCCAAAG 4001
Qy 1292 uAspGlyArgSerSerSerGlyProProHisGlyuThrAlaAlaProLyAspArgThrTrpAs 1312
Db 4002 GAGCGCAGAAAGCAGCTCAGAGACCCCGCATGAGACGGCGCGCCCAAGCGCACATAGA 4061
Qy 1312 pMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGlnGlyLeuMetG 1332
Db 4062 CATGATGAGAGGCGCGCTGGGACAGGCATCTCTCAAGCATCAAGAGTCTCATGGG 4121
Qy 1332 YATGAlaIleProProGlyuArgHisSerProHisAlaLeuLySgIleGlnHisIleAs 1352
Db 4122 CCGTGCATCCCGCGGAGGACACAGCCCAACCTCAAGAGAGACACCACTCCG 4181
Qy 1352 gGlySerIleThrGlnGlyIleProArgSerTrpValGluAlaGlnGlyuAspTrpLeuAr 1372

Db 4182 CCGGCTCATCAACAAGGATCCCTCGTCTACGTGAGGACACAGAGGACCTACCTCG 4241
Qy 1372 gATGAlaAlaLyLeuLeuLyAspArgGluGlyThrProProProProProSerArgAs 1392
Db 4242 TCGGAGGCGCAAGCTCTTAAGCGGAGGACAGCTCTCCGCCCCACCGCTTCACGGA 4301
Qy 1392 pLeuThrGluAlaTrpLySerThrGlnAlaLeuGlyProLeuLyLeuLyAspProAlaHisG 1412
Db 4302 CTGACCGGAGGCTCAACAAGACGAGGCTTGAGGCCCTGTAAGGTGAAGCGGCGCATAGA 4361
Qy 1412 uGlyLeuValAlaThrValLySgIleAlaGlyArgSerIleHisGlnIleProArgGlu 1432
Db 4362 GGGCTGTGTGCGCACGGTGAAGAGGGCGGCTCCATCAAGATCCCGCGGAGGA 4421
Qy 1432 uLeuArgHisThrProGlyuLeuProLeuAlaProArgProLeuLySgIleSerIleThr 1452
Db 4422 GCTGCGGACACCGCCGAGCTGCGCTGAGCCCGCGCGGCTCAAGAGGAGCTTCATCAC 4481
Qy 1452 rGlnGlyThrProLeuLySerTrpAspThrGlyAlaSerThrThrGlySerLySlyHisAs 1472
Db 4482 GAGGGGACCCCGCTCAAGTACGACCGGCGGTCCACCACTGGCTCCAAAAGACAGGA 4541
Qy 1472 pValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspVa 1492
Db 4542 CATAAGCTCCCTCATCGGACCGCCCGCGAGAGTTCCTCCACCGGTGACCCGCTGATGT 4601
Qy 1492 lMetAlaAspAlaArgAlaLeuGluArgGlyuArgGlyGlyuSerLeuLySerArgPr 1512
Db 4602 GATGCGGAGCGCGGCGACTGGAAGTGCTGCTCAAGAGAGAGCTGAAGAGCGCGCC 4661
Qy 1512 oGlyThrAlaSerSerSerGlyLySerIleAlaArgGlyAlaProValIleValProG 1532
Db 4662 AGGACCGGACGAGCTCGGGGGGCTCCATTTGGCGGGCGCCCGGTATTGGCTGA 4721
Qy 1532 uLeuGlyLySProArgGlySerProLeuThrTrpGlyuAspHisGlyAlaProPheAlaG 1552
Db 4722 GCTGGGTAAAGCCCGGACAGGCCCTGACCTTAGAGAACAGAGGACACCTTTGGCGG 4781
Qy 1552 YHisLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGlu 1572
Db 4782 CACCTTCCACAGAGTTCGCGCTGTACATGCGGAGCCACCGCGCTGACGAGAGG 4841
Qy 1572 ySerLeuSerSerSerLySlyAlaSerGlnAspArgLyLeuThrSerThrProArgGlu 1592
Db 4842 CAGCTTTGTTCAGGACAGGCATCCAGAGACGAAAGCTGAGCTGCGCTGTGATGAT 4901
Qy 1592 eAlaLySerProHisSerThrValProGlyuHisAspProHisProLleSerProTrpG 1612
Db 4902 CGCCAAAGTCCCGCACAGACCGTGCCTCGGACACCAACCCCATCTCGCCCTATGA 4961
Qy 1612 uHisLeuLeuArgGlyValSerGlyValaAspLeuTrpArgSerHisAlaProLeuAlaPh 1632
Db 4962 GACCTGCTTCGGGGGTGAAGTGCATGACCTGTATCGGAGCACATCCCTCGGCTT 5021
Qy 1632 eAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaLyTrpLeuPr 1652
Db 5022 CAGCCCACTCATACCCCGGGGATCCCTTGAGAGGAGCGGCTGCTACATACGCGC 5081
Qy 1652 oArgHisLeuAlaProAsnProThrTrpProHisLeuTrpProProTrpLeuLleArgG 1672
Db 5082 CCGACACCTGGGCGCCCAACCCCACTTACCGGACCTGTACCCACCTTACCTCCGCG 5141
Qy 1672 YTrpProAspThrAlaAlaLeuGlyuAsnArgGlnThrIleIleAsnAspTrpTrlIleThrSe 1692
Db 5142 CTACCCCGACAGCGCGCGCTGAGAAACGGGACAGACCATCATCAATGATCACTACCTC 5201
Qy 1692 rGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgG 1712
Db 5202 GAGACGATGACACACACACCGGACCGCATGCGCCAGCGAGTATGTTGAGGGG 5261
Qy 1712 YLeuSerProArgGlySerSerLeuAlaLeuAsnTrpAlaAlaGlyProArgGlyIleIle 1732
Db 5262 CCTCTGCGCCCGGAGTCTCTGTGGCACTCAATGCTGGGGGTCCCGAGGACATCAT 5321

QY 1732 eAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAl 1752
| | | | |
Db 5322 CGACGCTGCCAAGTCCACACCTGCTGCTGCGGCCCGGACACAGGAGCCCGACG 5381
QY 1752 aThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgH 1772
| | | | |
Db 5382 CACCGCATGAGACCGCTTGCTTACCTCCACCGCGGCCCGGACCTTGACGCGGCCCA 5441
QY 1772 sSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrSe 1792
| | | | |
Db 5442 CAGCACTCCCACTCTCCCGAGAGGTCCACACCTTGACAAACCAACCAACGCTC 5501
QY 1792 rSerSerGluArgGluArgAspArgAspArgGluArgAspArgGluArgGluArg 1812
| | | | |
Db 5502 CTCTGTCGAGCGGAGCGAGACCGGGATGAGACCGGGACCGGGATCGGGAGCGGAAA 5561
QY 1812 sSerLLeuThrSerThrThrValGluHisAlaProLLeuTyrArgProGlyThrG 1832
| | | | |
Db 5562 GTCCATCTTCACGTCACACGAGCGTGAGCAGCACCCATCTGAGAGACTGTGACAG 5621
QY 1832 uGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgProAl 1852
| | | | |
Db 5622 GCAGACGAGCGGAGCAGCGGAGCGGCGGGGTGGGAGCAGCAGCGCGCCGCG 5681
QY 1852 aSerHisSerHisAlaHisGlnHisSerProLLeuSerProArgThrGlnAspAlaLeuG 1872
| | | | |
Db 5682 CTCCCATCTCCCAAGCCCAACAGCACCTGCGCATCTCCCTCGGACCCAGGATCCCTCCA 5741
QY 1872 ngInaArgProSerValLeuHisAsnThrGlyMetLysGlyLLeuThrAlaValGluP 1892
| | | | |
Db 5742 GCAGAGACCCAGGTGCTTCAACAACAGGATGAAAGGATATCATCACCTCTGAGAGCC 5801
QY 1892 oSerLysProThrValLeuArgSerThrSerThrSerSerProValArgProAlaIaTh 1912
| | | | |
Db 5802 CAGCAAGCCCAAGCTCTGAGGTCACCTCACCTCTCCACCGGTGCGCAAGCTGAC 5861
QY 1912 rPheProProAlaThrHisGlySerProLeuGlyGlyThrLeuAspGlyValTyrProThrLe 1932
| | | | |
Db 5862 ATTCCACCTCCCAACCCACCACTGCGGCGGACCTCTGATGGGTCTAACCTTACCT 5921
QY 1932 uMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProAr 1952
| | | | |
Db 5922 CATGAGACCGGTCTTGCTGCTCCCAAGAGAGCGCCCGGCTGCGCGGCAAGCGCGCCG 5981
QY 1952 gAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAl 1972
| | | | |
Db 5982 AGCAGACACCGGCATGCTTCTCTGCGCAAGCCCCCGGCTCCGCGCTGAGCGCCGC 6041
QY 1972 aSerSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAl 1992
| | | | |
Db 6042 CTCTCTCCCGGAGAGGGCTCGAGGCCCGGCCCTGATGCTCTCTGTCGCGCACGC 6101
QY 1992 aThrLLeuAlaArgThrProAlaLysAsnLeuAlaProHisAlaLeuSerProAspProP 2012
| | | | |
Db 6102 CACCATGCGCGGACCGCTCGGAAGAACTCGGACCTCACGACCGCGGACCGCGCC 6161
QY 2012 oAlaProProAlaSerAlaSerAspProHisAlaArgGluLysThrGlnSerLysProPheSe 2032
| | | | |
Db 6162 GGGGCGCACCTGCTCGGCTCGGACCGGACCGGAAAGAACTCAAGTAAACCTTTTC 6221
QY 2032 rLLeuGlnLLeuGluLeuLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluG 2052
| | | | |
Db 6222 CATTCAGAACTGGAACTCCGTTCTCTGGGTACCAAGGACGACGACGACCCCGAAAG 6281
QY 2052 yValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLy 2072
| | | | |
Db 6282 GGTGAGACCGGTGACCGCTGAGCTCACCCAGTCTGACCAAGAGAGGCGTCCCCAA 6341
QY 2072 sHisLeuGlnGluLeuAspLysSerHisLeuGlnGlyGlyLeuLeuArgProLysGlnProG 2092
| | | | |
Db 6342 GCACCTTGAAAGAGCTCGAAGAGGACCTGAGAGGGAGGCTGCGGCCCAAGGAGCCAGG 6401

QY 2092 yProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuLeuArgProLeuProGluSe 2112
| | | | |
Db 6402 CCCCGTGAAGCTTGGCGGGAGGCGGCCCACTCCACACCTGCGGCGCTGAGAGG 6461
QY 2112 rGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgVa 2132
| | | | |
Db 6462 CAGGCCCTGCTCAGCCGCTCTCCAGACCGCCCGAGGGGTCCAAAGTCACACCGGCT 6521
QY 2132 lValThrLeuAlaGlnHisAlaLeuSerGlyValLLeuThrGlnAspTyrThrArgHisP 2152
| | | | |
Db 6522 GGTCAACCTCGGCCAGACATCATGAGGTATCACACAGACTACACCGGACACACCC 6581
QY 2152 oGlnGlnLeuSerAlaProLeuProAlaProLeuLysSerPheProGlyAlaLeuSerCysPr 2172
| | | | |
Db 6582 ACAGCAGCTCACCGGACCCCTGCGCCCGCTCTTACCTCTTCCCTGGGGCCAGCTGCCC 6641
QY 2172 oValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHisG 2192
| | | | |
Db 6642 GGTCTGAGCTTCGCGCGGCCCAACGATGACCTTACCTCCGCGCCCGGACCATGTGTC 6701
QY 2192 aProAlaArgGlySerProHisSerGlyGlyGlyLysArgSerProGluProAsnLysTh 2212
| | | | |
Db 6702 CCGGCGCGGTGCTCCCGCCACAGGAGGAGGCAAGAGTCTCCAGAGCCAAACAGAC 6761
QY 2212 rSerValLeuGlyGlyGlyGluAspGlyLLeuGluProValLysSerProProGluGlyMetTh 2232
| | | | |
Db 6762 GTCGGTCTTGGGGTGTGTGAGAGCGGTATTAACTGTGTCTCCACCGGAGGCGATGAC 6821
QY 2232 rGluProProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGlyGlnTh 2252
| | | | |
Db 6822 GAGGCCAGGCGACTCCCGAGTGTGTGTAACCGCTGCTGTAACGGGATGGGGAACAGAC 6881
QY 2252 rGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePh 2272
| | | | |
Db 6882 GAGGCCAAGATGAGTGGCTCCAAAGTCTCAGAGCAACACAGCCAGCCGACTTCTT 6941
QY 2272 eSerLysLeuThrGlnLysSerAsnSerAlaMetValLysSerLysLysGlnLLeuSny 2292
| | | | |
Db 6942 CAGCAAGCTGAGCGGAGGACCTCCGCTCATGTGTCAAGTCCAAAGAACAGATCAACA 7001
QY 2292 sLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnLLeuSerGlnProGlyThrG 2312
| | | | |
Db 7002 GAAGCTGAACACCCACCAACCGGAAATGAGCTGAAATTCATATTCAGCCAGCTGGAGGA 7061
QY 2312 uLLeuPheAsnMetProAlaLLeuThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaVa 2332
| | | | |
Db 7062 GATCTTCATATGCGCGGCATACCGGAACAGGCTTATGACTTATAGAAAGCCAGCGGT 7121
QY 2332 lGlnGluHisAlaLeuSerThrAsnMetGlyLeuGluAlaLLeuArgLysAlaLeuMetG 2352
| | | | |
Db 7122 GCAGAAACATGCGCAGACCAACATGGGGCTGGAGGCCATTAATTAAGAAAGCACTATGG 7181
QY 2352 yLysTyrAspGlnTnTnGluGlnLysSerProProLeuSerAlaAsnAlaPheAsnProLeuAs 2372
| | | | |
Db 7182 TAAATATGACCACTGGGAAAGAGTCCCGCGCTCAGCGGCAATGCTTTTAACTCTTGA 7241
QY 2372 nAlaSerAlaSerLeuProAlaAlaMetProLLeuThrAlaAlaAspGlyArgSerAspH 2392
| | | | |
Db 7242 TGCCATGCGACGCTCCGCTGCTATGCCATTAACGCTGTGAGCAGCAGAGGTACCA 7301
QY 2392 sThrLeuThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerAr 2412
| | | | |
Db 7302 CACACTCACCTCGCAGGTGCGCGGGAAGGCAAGGTCTCTGGGAGAACCCAGCGCCG 7361
QY 2412 gLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSe 2432
| | | | |
Db 7362 AAAAGCAAGTCCCGGCGCCCGGCTGTGAGTGGGAGACGGCCACCTCTCTCTCTC 7421
QY 2432 rValHisSerGlnGlyAspCysAsnArgArgThrProLeuThrAsnArgValTyrGluAs 2452
| | | | |
Db 7422 AGTGCATCTGGAGGAGACTGCAACCGCGGAGCGGCTTACCAACCGGTGTGGAGGA 7481
QY 2452 pArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuLLeuMetArgLeuG 2472

Db	7492	CAGGCGCTCGTCCGAGGTTCCAGCCATTCCCTTCAACCCCTGATCATGCGGCTGCA	75411
Oy	2472	nalaglyValMetalaSerProProProglyLeuProAlaGlySerGlyProLeuAl	24922
Db	7542	GCGGGGTGTCAATGAGTCTCCCAACCCCAACCGGCGCTCCCGGGGCAAGCGGCGCTCGC	7601
Oy	2492	AGLYProHisAlaATPAspGluGluProLyAspProLeuLeuCySserGlnTYrGluWth	2512
Db	7602	TGGCCCCCACACACGCGCTGGAGAGAGAGCCCAAGCCACTGCTGCTCGAGTACGAGAC	7661
Oy	2512	rLeuSerAspSerGlu 2517	
Db	7662	ACTCTCGACAGCGAG 7677	
RESULT 12			
ID	ADG86290	standard; cDNA; 8686 BP.	
XX	AC	ADG86290;	
XX	DT	11-MAR-2004 (first entry)	
XX	DE	Human SMRT encoding cDNA SEQ ID NO:4.	
XX	KW	SMRT; silencing mediator for retinoid and thyroid hormone action;	
KW	KW	SMRT inhibitor; cytosolic; antiinflammatory; antiarthritic;	
KW	KW	antirheumatic; antisense therapy; inflammatory disorder;	
KW	KW	rheumatoid arthritis; hyperproliferative disorder; cancer; leukaemia;	
KW	KW	breast cancer; human; gene; ss.	
XX	OS	Homo sapiens.	
XX	FT	Key	Location/Qualifiers
FT	CDS	157..7680	
FT	FT	/*tag= a	
FT	FT	/product= "SMRT"	
XX	XX	MO2003106645-A2.	
XX	XX	24-DEC-2003.	
XX	XX	17-JUN-2003; 2003WO-US018923.	
XX	XX	17-JUN-2002; 2002US-00174014.	
XX	PA	(ISIS-) ISIS PHARM INC.	
XX	PI	Bennett CF, Freiler SM, Doble KW;	
XX	PI	WPI: 2004-082184/08.	
DR	DR	P-PSDB; ADG86291.	
DR	DR	GENBANK; AF125672.	
XX	XX	Novel antisense compound targeted to nucleic acid encoding SMRT	
PT	PT	(silencing mediator for retinoid and thyroid hormone action), useful for	
PT	PT	treating animal having disease associated with SMRT such as cancer,	
PT	PT	rheumatoid arthritis.	
XX	XX	Example 13; SEQ ID NO 4; 260bp; English.	
XX	XX	The present invention describes a compound (I) 8-50 nucleobases in length	
CC	CC	targeted to a nucleic acid molecule encoding SMRT (silencing mediator for	
CC	CC	retinoid and thyroid hormone action), where (I) specifically hybridises	
CC	CC	with the nucleic acid molecule encoding SMRT and inhibits expression of	
CC	CC	SMRT. (I) specifically hybridises with at least 8-nucleobase portion of a	
CC	CC	preferred target region on nucleic acid molecule encoding SMRT. Also	
CC	CC	described is a composition (II) comprising (I) and a carrier or diluent.	
CC	CC	(I) and (II) have cytostatic, antiinflammatory, antiarthritic and	
CC	CC	antirheumatic activities, and can be used in antisense therapy, and as	
CC	CC	SMRT expression inhibitors. (I) is useful for inhibiting the expression	
CC	CC	of SMRT in cells or tissues. (I) is also useful for treating an animal	
CC	CC	having a disease or condition associated with SMRT, e.g., inflammatory	

	disorder such as rheumatoid arthritis; or a hyperproliferative disorder
CC	such as cancer chosen from leukaemia and breast cancer, by inhibiting the
CC	expression of SMRT. (I) is useful for diagnostics, therapeutics,
CC	prophylaxis and as research reagents and kits. The present sequence
CC	encodes human SMRT, which is used in an example from the present
CC	invention.
XX	
SQ	Sequence 8686 BP; 1861 A; 3069 C; 2562 G; 1174 T; 0 U; 0 Other;
 Alignment Scores:	
Pred. No.:	0 Length: 8686
Score:	12978.00 Matches: 2487
Percent Similarity:	98.57% Conservative: 3
Best Local Similarity:	98.46% Mismatches: 9
Query Match:	98.21% Indels: 28
DB:	12 Gaps: 3
 US-09-522-753-5 (1-2517) x ADG86290 (1-8686)	
Dy	1 MetSerGlySerThrGlnLeuValAlaGlnIleThrPArgAlaThrGluProArgTyrPro 20
Db	157 ATGTGGAGCTCCACAGCGCTGTGGACAAGATGGAGGGCACTAGGCCCTCCTACC 216
Dy	21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
Db	217 CCCACACACTTTCTTACCCTGCAGATCGCCGACCCACACAGGACGTCGGGCTCTTG 276
Dy	41 GlyTyrGlnHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
Db	277 GAGTACCAAGCACACTCCCGGACATAGCTCTCCACTTCGCCGGCTCATCATCCAG 336
Dy	61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyValangLuarGserGln 80
Db	337 CCCACGGGGCGGAGGCCCTCCCTGCTGTGAATTCCAGCCCGGAAATGACGCTCCAG 396
Dy	81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet 100
Db	397 GAGCTCCAACCTCGCGCCAGAGTCCCACATCATCTCGGCCGAGACTGGGGAAATGCAGAGATG 456
Dy	101 GluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuAAspPro 120
Db	457 GAGTTCATTGAAGAAGCAAGCGCCCTCGGCTAGAGCTGCTCCTGACCCCTGCTGCACCG 516
Dy	121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
Db	517 TCACCCCTCGCTGGCCACCGGCCAGGCTGGGGATCTGAAGAAGCTCCACCAAGAACCGTAGC 576
Dy	141 LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
Db	577 CTGAAGGGCAAGCTGAACCGGATGCTCCCCCAGACCCCCCAGCACATAACCTTGAGCTG 636
Dy	161 GluLeuValProProArgLeuSerLysGluLeuIleGlnAsnMetLaspArgValAsp 180
Db	637 GAGCTGGTGGCCGCCACCGGCTGTCCAAAGAGGAGCTGATCCAAACATGGACCGCGTGAC 696
Dy	181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeu 200
Db	697 CGAGGATCACCATGTATGAGACAGAGATCTTAAGCTGAABAAGAAGACAGCAACGCTG 756
Dy	201 GluGluGlnAlaAlaLysProProGluProGluLysProValSerProProIleGlu 220
Db	757 GAGGAGGAGGCTGCCAACCGCCGCCAGGCTGAGAAAGCCGCTGTACCGCCGCCCATCGAG 816
Dy	221 SerLysHisLaspSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAla 240
Db	817 TCGAAGACCCGACGCTGTGTGACATCTTACGACGAAACCGGAAGAAGCTGAAGCT 876
Dy	241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrArgGlnPro 260
Db	877 GCACATCGGATTTCTGAAAGGCTGGGGCCCGAGCTGTGAGACTCCCTGTACAAACAGCCC 936
Dy	261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280

Db 937 TCCGACACCCGGCAGTATCATGAGAAATCAATAAATTAACCGCGATGCGGAGAAAGCTA 996
Qy 281 ILeuLeuTyrPheLeuYsrghArgshnHsaiAargLyseGlnTrpLyseGlnYsrPheYseGln 300
Db 997 ATCTTGTACTTCAAGAGAGAGAAATCCGCTCCGAAACAATGGAGAGCAAGATTCTCCAG 1056
Qy 301 ArgTyrAspGlnLeuMetGlnAlaLeuGlnYlyValGlnArgGlnLeuAsnPro 320
Db 1057 CCTATGACACAGCTCATGAGAGCTTGAGAGAAAGTAGAGAGCGATCCAGAAACAACCC 1116
Qy 321 ArgArgArgAlaAlaLeuLeuSerLyseValArgGlnTyrTyrGlnYlyseGlnPheProGln 340
Db 1117 CCGCGCGCGGCGCAAGAGAGCAAGGTTCCGAGTACTACAGAGAGCAAGTTCCCTGAGATC 1176
Qy 341 ArgLyseGlnArgGlnLeuGlnGlnArgMetGlnSerArgValGlnYlnArgLyseGln 360
Db 1177 CCGACAGACCGCAGCTGCAGAGAGCGCATGCGAG--AGGGTGGGCCAGCGCGGCACTGG 1233
Qy 361 LeuSerMetSerAlaAlaArgSerGlnYlyseGlnValSerGlnYlyleAspGlyLeuSer 380
Db 1234 CTGTCCATGTCCGCCCGCCGAGCGAGCAAGAGTGTCAAGATCATCATGAGCTCTCA 1293
Qy 381 GlnGlnGlnYlyAsnLeuGlnYlyseGlnMetArgGlnLeuAlaValIleProPheMetLeuTyr 400
Db 1294 GAGCAGAGAACCTGGAGAGAGAGATGCGCAGCTGCGGTGATCCCGCCCATGCTGAC 1353
Qy 401 AspAlaAspGlnGlnArgIleYsrPheIleAsnMetAsnGlyLeuMetAlaAspPheMet 420
Db 1354 GACGCTGACACAGAGCGCATCAAGTTCACTACATGAACGCGCTTATGGCCGACCCCATG 1413
Qy 421 YlyValYlyLyAspArgGlnValMetAsnMetTrpSerGlnGlnGlnYlyseGlnThrPhe 440
Db 1414 AAGGTGTACAAAGACCGCCAGGTCATGAACATGTGAGTAGAGAGAGAGAGAGACCTTC 1473
Qy 441 ArgGlnYlyPheMetGlnHsIleProYlyAsnPheGlyLeuIleAsnSerPheLeuGlnArg 460
Db 1474 CCGGAGAAAGTTCAAGATGCAATCCCAAGAACTTTGGCTGATGCAATCTCTCGAGAGG 1533
Qy 461 YlyThrValAlaGlnCysValLeuTyrTyrTyrLeuThrYlyAsnGlnYlyAsnYlyLys 480
Db 1534 AAGACGTGGCTGAGTGGCTCTCTATTACTTACTGACTGAGAAAGATGAACTTAAG 1593
Qy 481 SerLeuValArgArgSerTyrArgArgArgGlyLyseSerGlnGlnGlnGlnGlnGln 500
Db 1594 AGCTGTGTGAGAGAGAGCTATCGCGCCGCGGCAAGAGCAGCAGCAACAACAGCAGAG 1653
Qy 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
Db 1654 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1713
Qy 521 AspGlnYlyseGlnYlyseGlnYlyseGlnYlyseGlnYlyseGlnYlyseGlnYlyse 540
Db 1714 GATGAG 1773
Qy 541 AspLyseGlnYlyseGlnYlyseGlnYlyseGlnYlyseGlnYlyseGlnYlyseGln 560
Db 1774 GACAG 1833
Qy 561 YlyseGlnYlyseGlnYlyseGlnYlyseGlnYlyseGlnYlyseGlnYlyseGln 580
Db 1834 AAG 1893
Qy 581 ArgIleThrArgSerMetAlaAsnGlnYlyseGlnYlyseGlnYlyseGlnYlyseGln 600
Db 1894 CGCATCACCGCTCAATGGCTTAATGAGGCCAACAAGAGAGAGAGAGAGAGAGAGAG 1953
Qy 601 SerAlaGlnLeuAlaSerMetGlnYlyseGlnYlyseGlnYlyseGlnYlyseGln 620
Db 1954 AGCGCGAG 2013
Qy 621 GlnThrAlaYlyseGlnYlyseGlnYlyseGlnYlyseGlnYlyseGlnYlyseGln 640
Db 2014 GAAACAGCCCAAGAAAGGTCTCTTGAAACAAGCGCGCAACTGCGGCAATGCGCCGAG 2073

Qy 641 ValGlySerLyseThrValSerGlnCysYlyAsnPheTyrPheAsnTyrYlyseArgGln 660
Db 2074 GTGGGCTCCAAAGACTGTGTCGAGTGTAAAGAACTTCTTCACTCAACAAGAAAGAGCAG 2133
Qy 661 AsnLeuAspGlnIleLeuGlnGlnHsIlyYlyseYlyseMetGlnYlyseGlnYlyseGln 680
Db 2134 AACCTTGATGAGATCTTGAGAGAGCAAGCTGAAAGTAGAGAGAGAGAGAGAGAGAG 2193
Qy 681 ArgYlyLyseYlyseAlaProAlaAlaAsnSerGlnYlyseGlnYlyseGlnYlyseGln 700
Db 2194 AGGAAG 2253
Qy 701 GlnAspGlnGlnMetGlnYlyseGlnYlyseGlnYlyseGlnYlyseGlnYlyseGln 720
Db 2254 GAGGATGAGAGATGAG 2313
Qy 721 AlaGlnYlyseYlyseAlaSerGlnYlyseGlnYlyseGlnYlyseGlnYlyseGln 740
Db 2314 GCTGAA-----GCC 2322
Qy 741 ThrValAsnAsnSerSerAspThrGlnSerIleProSerProHsIleThrGlnYlyseGln 760
Db 2323 ACTGTCAACACAGCTCAGACACCGAGAGCATCCCTCTCTCACTGAGGCGCGCAAG 2382
Qy 761 AspThrGlnYlyseGlnYlyseGlnYlyseGlnYlyseGlnYlyseGlnYlyseGln 780
Db 2383 GACACAGGCGCAAGATGGCGCCAAAGCCCGCAGCCCTGGCGCGAGCGGCGCACCCCA 2442
Qy 781 GlnProProProProProProProProProProProProProProProProProPro 800
Db 2443 GGGCCACCCACCCACCCACCCAGAGAGATCCCGGCC-CCACTGAGTCCACCCCGGCTTC 2501
Qy 800 GlnYlyleThrGlnYlyleProProProProProProProProProProProProPro 820
Db 2502 TGAAGCCACTTAAGCCCTACAGCCCGCCACACAGACCCCATTTCCCTCTTCACTCTTC 2561
Qy 820 OValYlyProYlyseGlnYlyseGlnYlyseGlnYlyseGlnYlyseGlnYlyseGln 840
Db 2562 TGTGTCCCAAG 2621
Qy 840 YlyGlnGlnYlyseProProAlaAlaGlnYlyseGlnYlyseGlnYlyseGlnYlyseGln 860
Db 2622 GGAG 2681
Qy 860 UProValYlyseGlnYlyseGlnYlyseGlnYlyseGlnYlyseGlnYlyseGln 880
Db 2682 GCCCTCAAG 2741
Qy 880 AglnYlyleGlnYlyleThrAlaGlnYlyleYlyleYlyleYlyleYlyleYlyse 900
Db 2742 GGAAGCGCTGAGGCGCACGCGCGAGAGGCGCTCAAGGCGAGAGAGAGAGAGAGAG 2801
Qy 900 GlnYlyleGlnYlyleThrAlaYlyseSerGlnYlyleProGlnYlyleAspSerSerAla 920
Db 2802 CCGCAGGCGCACAG 2861
Qy 920 YlyseSerAlaAspGlnYlyleYlyleYlyleYlyleYlyleYlyleYlyleYlyleYly 940
Db 2862 CTGCACTGAG 2921
Qy 940 OArgProSerLeuLeuThrProThrGlnYlyleProAlaAlaAsnYlyleSerProGlnYly 960
Db 2922 AAGGCCACAGCTCTCAACCCGAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2981
Qy 960 OLeuAspLeuYlyseGlnYlyseGlnYlyleYlyleYlyleYlyleYlyleYlyleYly 980
Db 2982 ACTGGAAG 3041
Qy 980 GlnYlyleGlnYlyleProProAlaAlaProThrYlyleProProAlaProProAlaPro 1000
Db 3042 AGTCAATGAG 3101

QY 1000 oProProGlnaLnleuGlnProGlnuSerAspAlaProGlnGlnProGlnSerSerProAr 1020
DB 3102 GCGACCCCAAAACCTGACGCGGAGAGGAGGAGCCCTCAGACGCTGAGACAGACCCCG 3161
QY 1020 gGlyLysSerThrsSerProAlaProProAlaAspLysGlu----- 1033
DB 3162 GGGGAAAGACAGAGGAGCCGCGACCCCGCGACAGAGAGGACAGAGACCTGTGTTCTT 3221
QY 1034 -----AlaPheAlaAlaGlnAlaGlnLysLeuProGlnYAspProProCysTrpThrsSerG 1052
DB 3222 CCGAGCCTTGCGAGCCGACGAGCCGAGAGAGCTGCTGGGAGACCCCTTGCTGAGACTTCGG 3281
QY 1052 YLeuProPheProValProProArgGlnValLLeuAlaSerProHisAlaProAspPr 1072
DB 3282 CTTGCTCCTTCCCGTGCCCGCGTGAGGTATCAAGGCGTCCCGCATGCCCCGAGACC 3341
QY 1072 oSerAlaPheSerThrsAlaProProGlnYHisProLeuProLeuGlnYLeuHisAspThrAl 1092
DB 3342 CTCAGCCTTCTCTACACCTCCACCTGGTCAACCACTGCCCCCTGGGCTTCATGACACTGC 3401
QY 1092 aArgProValLeuProArgProProThrLLeuSerAsnProProProLeuLLeuSerAl 1112
DB 3402 CCGGCGGCTGCTGCGGCGGACCCACCATCTCCAACTCCGCTCCCTCATCTCTCTGC 3461
QY 1112 aLYsHisProSerValLeuGlnArgGlnLLeuGlnAlaLLeuSerGlnLysMetSerValG 1132
DB 3462 CAAGCAACCCAGACGCTCTCGAGAGGCAAAATAGTGCATCTCCCAAGAAATGTCGTTCA 3521
QY 1132 nLeuHisValProTrpSerGlnHisAlaLysAlaProValGlyProValThrMetGlyLe 1152
DB 3522 GCTTCAAGTCCCGTCACTCAAGAGCATGCCAAAGCCCGGTGGCCCTGTCAACATGGGGCT 3581
QY 1152 uProLeuProMetAspProLysLysLysAlaProPheSerGlyValLysGlnGlnLLeu 1172
DB 3582 GCGCTGCGCATGAGACCCCAAAAGCTGCGACCTTCAAGCGAATGAAGAGGAGCGCT 3641
QY 1172 uSerProArgGlyGlnAlaGlyProProGlnuSerLeuGlnYValProThrAlaGlnGlnAl 1192
DB 3642 GTCCCAACGGGGCCAGGCTGGGCGCACCGAGAGCTGGGGGTGCCCAACGCCAGAGAGC 3701
QY 1192 aSerValLeuArgGlyThrAlaLeuGlnYSerValProGlnYGlySerLLeuThrLysGlyL 1212
DB 3702 GTCCGCTGCTGAGAGGAGACACTCGGGCTCAGTTCCGGGCGAAGGATCAACAAAGGAT 3761
QY 1212 eProSerThrArgValProSerAspSerAlaLLeuThrTrpArgGlySerLLeuThrHisG 1232
DB 3762 TCCAGACACGAGGTGCTCTGAGACAGCGCATCACTACCGGCTCCATCACCCACGG 3821
QY 1232 YThrProAlaAspValLeuTrpLysGlyThrLLeuThrArgLLeuGlnYGlnuAspSerPr 1252
DB 3822 CAGGCCAGCTGAGCTCTGTHCAAGGAGCACATCAACAGATCATCGGCGAGAGACGCC 3881
QY 1252 oSerArgLeuAspArgGlyArgGlnuAspSerLeuProLysGlyHisValLLeuTrpGln 1272
DB 3882 GAGTGGCTTGACCGGCGCGGAGGAGACAGCTGCGCAAGGGCGACGTCAACGAAG 3941
QY 1272 YLYsLYsGlyHisValLeuSerTrpGlnGlnYGlyLysMetSerValThrGlnYSerLysG 1292
DB 3942 CAAGAAGGGGCAAGTCTTGTCTATGAGGGTGGCATCTCTGTGACCCAGAGCTCCCAAG 4001
QY 1292 uAspGlyArgSerSerSerGlyProProHisGlnuThrAlaAlaProLysArgTrpThrTrpAs 1312
DB 4002 GAGCGGAGAGAGAGCTCAAGAGACCCCGCATGAGACGAGCGCGCCCAAGGAGCATATGA 4061
QY 1312 pMetMetGlnGlyArgValGlyArgAlaLLeuSerSerAlaSerLLeuGlnYLeuMetG 1332
DB 4062 CATGATGAGAGCGCGCTGGGACAGACCATCTCTCAGCCAGCATCAAGAGTCTCATGGG 4121
QY 1332 YArgAlaLLeuProProGlnuArgHisSerProHisLysLeuLysGlnGlnHisGlnLLeuAr 1352
DB 4122 CCGTGCCATCCCGCGAGGAGACACAGCCCGCACCTCAAAAGAGAGACCAACATCCG 4181
QY 1352 gGlySerLLeuThrGlnGlnLysLLeuProArgSerTrpValGlnAlaGlnGlnuAspTrpLLeuAr 1372

DB 4182 CCGGCTCATGACACAGGAGATCCCTCGGTCTCATGTGAGAGGACAGAGGATCTACGCG 4241
QY 1372 gArgGlnAlaLysLeuLeuLysArgGlnGlnYThrProProProProProSerArgAs 1392
DB 4242 TCGGAGAGGCAAGTCTTAAAGCGGAGGAGGACCTTCGCGCCACCGCGCTCAAGGGA 4301
QY 1392 pLeuThrGlnAlaLysLysThrGlnAlaLeuGlyProLeuLysLysLeuLysProAlaHisG 1412
DB 4302 CTGACCGAGAGCTTCAAGACAGAGCGCTGCGGCCCTGTAAGCTGAAGCGGCGCATGA 4361
QY 1412 uGlyLysValAlaThrValLysGlnAlaGlyArgSerLHisGlnLLeuProArgGln 1432
DB 4362 GGGCTGTGTGGCCACGGTGAAAGAGGCGGCGCCCTCATCATGATCCCGCGGAGGA 4421
QY 1432 uLeuArgHisLLeuProGlnuLeuProLeuAlaProArgProLeuLysGlnLysLLeuTh 1452
DB 4422 GTGGCGGACACCGCCAGCTGCGCTGCGCGCGCGCGCTCAAGAGAGGAGCTCATCAC 4481
QY 1452 rGlnGlyThrProLeuLysTrpAspThrGlyAlaSerThrGlySerLysLysHisAs 1472
DB 4482 GAGAGGACACCCGCTCAAGTACAGACCGGCGGTCCACACCTGCTCAAAAGACGA 4541
QY 1472 pValArgSerLeuLLeuYSerProGlyArgThrPheProProValHisProLeuAspVa 1492
DB 4542 CGTACGCTCCTCATGGGAGCCCGGCGGAGCTTCCACCGTGACCCGCTGATGT 4601
QY 1492 lMetAlaAspAlaArgAlaLeuGlnuArgLAcGlyTrpGlnuSerLeuLysSerArgPr 1512
DB 4602 GATGGCGAGCGCGGCGCATGGAACTGCTCTCTCAAGAGAGCGCTGAAGACCCGCG 4661
QY 1512 oGlyThrAlaSerSerSerGlyLysSerLLeuAlaArgGlyAlaProValLLeuValProG 1532
DB 4662 AGGAGCGGAGAGCTCGGAGGAGCTCATTTGCGCGGCGCGCGCTCATTTGTGCTGA 4721
QY 1532 uLeuGlyLysProArgGlnSerProLeuThrTrpGlnuAspHisGlyAlaProPheAlaG 1552
DB 4722 GCTGGGTAAACCGCGGAGAGCCCGCTGACTATGAGGACACAGGGGCACTTTGCGCG 4781
QY 1552 YHisLeuProArgGlyLysSerProValThrMetArgGlnProThrProArgLeuGlnGln 1572
DB 4782 CCACTTCCCAAGAGTTCGCGCGGTGACCATCGGAGGCCACCGCGCTGCGAGGAGG 4841
QY 1572 YSerLeuSerSerSerLysValaSerGlnaAspArgLysLLeuThrSerThrProArgGln 1592
DB 4842 CAGCCTTCTGTCAGAGAGGATCCAGAGCCAAAGCTGACGTGAGCCTGTGATGAT 4901
QY 1592 eAlaLysSerProHisSerThrValProGlnuHisLAspProLLeuLLeuSerProTrpG 1612
DB 4902 CGCAGAGTCCCGGACAGACACCGTGCAGAGACACACCCACATCTCGGCTATGA 4961
QY 1612 uHisLeuLeuArgGlyValaSerGlyValaLeuLeuTrpArgSerHisLLeuProLeuAlaP 1632
DB 4962 GCACCTGCTTCGGGGGTGAGTGGCGTGAACCTGTATCCGACACATCCCTGCGCTT 5021
QY 1632 eAspProThrSerLLeuProArgGlyLysLLeuAspAlaAlaAlaLysTrpLLeuPr 1652
DB 5022 CAGACCCCATCTCAACCCCGCGGAGTCCCTGAGAGCGCGCTGCTTACATCCTGCC 5081
QY 1652 oArgHisLLeuAlaProAspProThrTrpProHisLLeuTrpProProTrpLLeuArg 1672
DB 5082 CCGACACCTGCGCCCGCAACCCCATACCGGACCTGTACCCACCTACTCATCCGCGG 5141
QY 1672 YTrpProAspThrAlaAlaLeuGlnuAsnArgGlnuThrLLeuLAsnAspTrpLLeuThrs 1692
DB 5142 CTACCCCGACACGCGCGCTGAGAAACGGGACAGCATCATGATCAATCACTACCTC 5201
QY 1692 rGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnuArgAlaAspMetLeuArgG 1712
DB 5202 GAGAGAGATGACACACACGCGGACCGCATGGCGCGAGGAGTATGCTGAGAGGG 5261
QY 1712 YLeuSerProArgGlnuSerSerLeuAlaLeuAsnTrpAlaAlaGlyProArgGlyLLeuL 1732

Db 5262 CCTCTGCCCCGGAGTCTCTGGCACTCAACGCTGGGGTCCCCGAGGCATCAT 5321
Qy 1732 eAspLeuSerGlnValProHisleuProValIleuValProProThrProGlyThrProAl 1752
Db 5322 CCACTGTCCCAAGTCCACACTGTGCTGTGCTGTGCCCCGACACACGAGCCACAG 5381
Qy 1752 aThrAlaMetAspArgLeuAlaTyrIleuProThrAlaProGlnProPheSerArgHis 1772
Db 5382 CACCGCATGAGACCGCTTGCTTCACTCCCAACGGGCCCCAGCCCTTCAGACCGGCCA 5441
Qy 1772 sSerSerSerProLeuSerProGlyIleProThrHisleuThrTylsleuProThrThrIse 1792
Db 5442 CAGCACTCTCCCACTCTCCCAAGAGGTCCACACACTTGACAAACCAACCAACGCTC 5501
Qy 1792 rSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGlu 1812
Db 5502 CTCTCTCCGAGCGGAGACGAGACCGGATCGAGACCGGAGATCGGAGCGGAGAAA 5561
Qy 1812 sSerIleleuThrSerThrThrThrValGluHisAlaProIleTyrArgProGlyThrG 1832
Db 5562 GTCCATCTCTCACTCCACACGAGGTGAGACGCAACCATCTGAGAGACTGTACAGA 5621
Qy 1832 uGlnSerSerGlySerSerGlySerSerGlyIleGlyIleGlyIleSerSerSerArgProAl 1852
Db 5622 GCAAGACAGCGGAG 5681
Qy 1852 aSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuG 1872
Db 5682 CTCCCACTCCCACTCCCAACGACACTGCCCCATCTCCCTCGACACCGAGATCCCTCCA 5741
Qy 1872 ngInaArgProSerValIleuHisAsnThrGlyMetIleGlyIleIleThrAlaValGlnPr 1892
Db 5742 GCAGAGACCCAGAGTGTCTTCAACAACAGGATGAGAGATATCATCACCTGTGAGAGCC 5801
Qy 1892 oSerIleProThrValIleuArgSerThrSerThrSerSerProValArgProAlaIleThr 1912
Db 5802 CAGCAAGCCCAACGCTGTGAGGTCCACTCCACTCTTCAACCGTTCGCGACACTGCGCAC 5861
Qy 1912 rPheProProAlaThrHisAcYsProLeuGlyIleThrLeuAspGlyValTyrProThrIle 1932
Db 5862 ATTTCCCACTGCCACCACTGCCCACTGGGCGGACCTCGATGGGGTCTAACCTTACCT 5921
Qy 1932 uMetGluProValIleuLeuProIleuArgIleAlaProArgValAlaArgProGluArgProAr 1952
Db 5922 CATGGAGCCCGTCTTGTCTGCGCAAGAGGCCCCCGGGTGTGCGCGAGAGGCGCCCG 5981
Qy 1952 gAlaAspThrGlyHisAlaPheLeuAlaIleProProAlaArgSerGlyIleuGluProAl 1972
Db 5982 AGCAGACACCGGACATGCTTCTCTGCGCAAGCCCCAGCCGCTCCGGGCTGAGAGCCCGC 6041
Qy 1972 aSerSerProSerIleuSerGlyIleProArgProLeuValProProValSerGlyHisAla 1992
Db 6042 CTCTCTCCCCAGAGAGGCTCTGGAGCGCCGAGCCCTTAGTGTCTTCTGTGCGCACGC 6101
Qy 1992 aThrIleAlaArgThrProAlaIleuAsnLeuAlaProHisHisAlaSerProAspProPr 2012
Db 6102 CACCAATGCGCGGACCCCTGCGAAGAACTCGCACCTCACACGCGACCGCGAGCCCGCC 6161
Qy 2012 oAlaProProAlaSerAlaSerAspProHisArgGlyIleuThrGlnSerIleuProPhe 2032
Db 6162 GGGGCGCACTGCTCGGCTCTCGAGCCGCGACCGGAGAAAGACTCAAAAGTAAACCTTTTTC 6221
Qy 2032 rIleGlnIleuGluIleuArgSerIleuGlyTyrHisGlySerSerTyrSerProGlyIle 2052
Db 6222 CATTCAGAACTGAAGTCTGTTCTCTGGGTACACAGGAGAGAGACTTACAGCCCCGAAAG 6281
Qy 2052 yValGluProValSerProValSerSerProSerIleuThrHisAspIleuGlyLeuProLy 2072
Db 6282 GGTGGAGCCCGGTGAGCCCTGTGAGCTCACCACTGTAACCAAGAGAGAGGCTCCCCCAA 6341
Qy 2072 sHisIleuGluIleuAspIleuSerHisIleuGluGlyIleuLeuArgProIleuGlnProG 2092
Db 6342 GCACCTGGAAGAGCTGCAAGAGCCACTGAGAGGGAGAGCTGCGGCCAAGAGAGCAGG 6401

Qy 2092 yProValIleuGlyIleGluAlaAlaHisleuProHisleuArgProLeuProGluSe 2112
Db 6402 CCCCCTGAAGCTTGGCGGAGAGCGCCCACTCCACACTGCGGCGCTGCTGCTAGAG 6461
Qy 2112 rGlnProSerSerSerProLeuGlnThrAlaProGlyValIleuGlyHisGlnArgVa 2132
Db 6462 CAGGCTCTGTCCAGCCCTGCTCCAGACCGGCCAGGAGGTCAAAAGTCAACAGCGGGT 6521
Qy 2132 lValThrIleuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisIlePr 2152
Db 6522 GGTCACTCTGGCCGACATCATGAGGTATCATACAGAGACTTACACCGGACCAACCC 6581
Qy 2152 oGlnIleuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysPr 2172
Db 6582 ACAGCAGCTCAGGCGACCCCTGCGCCCGCCCTCTACTCTCTCCCTGGGGCCGCTGCCC 6641
Qy 2172 oValIleuAspLeuArgArgProProSerArgLeuTyrIleuProProProAspHisGlyAl 2192
Db 6642 CTCTCTGAGCTTCGCGGCCCAACGATACCTTACCTTCGCGCCCGGACCATGATGC 6701
Qy 2192 aProAlaArgGlySerProHisSerGlyIleGlyIleArgSerProGluProAsnIleuThr 2212
Db 6702 CCGGCGCCGTGCTCCCGCCACAGCGAAGGGGCAAGAGTCTCCAGAGCCAAACAAGAC 6761
Qy 2212 rSerValIleuGlyIleGlyIleuAspGlyIleGluProValSerProProGlyIleMetTh 2232
Db 6762 GTCGGTCTGGGTGTGTGAGGACCGTATTGAACTGTGTCCCAACCGGAGGGCATGAC 6821
Qy 2232 rGluProProGlyHisSerArgSerAlaValTyrProIleuLeuTyrArgAspGlyIleGlnThr 2252
Db 6822 GGAGCCAGGCGACCTCCCGAGTGTGTGTACCGCTGTGTACCGGAGTGGGAGAACAGAC 6881
Qy 2252 rGluProSerArgMetGlySerIleuSerProGlyIleAsnThrSerGlnProProAlaPhePh 2272
Db 6882 GGAAGCCAGAGATGGCTCCAGTCTCCAGGCAACACAGCCACCGCCACCTTCTT 6941
Qy 2272 eSerIleuThrGluSerAsnSerAlaMetValIleuSerIleuGlnIleuAsnIle 2292
Db 6942 CAGCAAGTGAACCGAAGACACTCCGCAATGTGCAAGTCCAGAAACAGAGATCAACAA 7001
Qy 2292 sIleuLeuAsnThrHisAsnArgAsnGluProGlyTyrAsnIleSerGlnProGlyThrG 2312
Db 7002 GAAGCTGAACACCAACACCGAATAGCCTGAAATCAATATCAGCAGGCTTGAGAGCGGA 7061
Qy 2312 uIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaVa 2332
Db 7062 GATCTTCAATATGCCCGCCATTCACCGGAACAGGCTTATGACTTATAGAAAGCACTATGGG 7121
Qy 2332 lGlnIleuHisAlaSerThrAsnMetGlyIleuGluAlaIleIleArgIleuAlaLeuMetG 2352
Db 7122 GCAAGAACTGCCAGACCAACCAATGGGGCTGGAGGCCATATATTAAGAAAGCACTATGGG 7181
Qy 2352 yIleTyrAspGlnTyrGluIleuSerProProLeuSerAlaAsnAlaPheAsnProLeuAs 2372
Db 7182 TAAATATGACCAAGTGGGAAGAGTCCCGCGCTGACGCGCAAAAGCTTTTAACCTCTGAA 7241
Qy 2372 nIleAsnAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHis 2392
Db 7242 TGCAGGTGCGAGCTCTCCGCTGTATGCGCATTAACCGTCTGAGAGAGAGGTATACCA 7301
Qy 2392 sThrIleuThrSerProGlyIleGlyIleuIleuAsnIleuValIleSerGlyArgProSerSerAr 2412
Db 7302 CACACTCACCTGCGCAGGTGGGGGGGAAAGGCAAGGTCTCTGGAGAGACCGAGAGCCG 7361
Qy 2412 gIleuAlaIleuSerProAlaProGlyIleuAlaSerGlyIleuAspArgProProSerValSerSe 2432
Db 7362 AAAAGCAAGTCCCGCGCCCGGCTGTGATCTGGGAGACCGGCGCAACCTCTGTCTCTC 7421
Qy 2432 rValHisSerGluGlyIleuAspCysAsnArgArgThrProLeuThrAsnArgValTyrGluAs 2452
Db 7422 AGTGACTGGAGAGAGATGCAACCGCGGAGAGCGGCTTCAACCAACCGCGTGTGGAGAGA 7481

QY 2452 pargProSerSerA1aGlySerThrProPheProTyAanProLeuIleMetArgLeuG1 2472
DB 7482 CAGGCGCTCGTCCGACGGTTCACCGCATTCCTCCACAGCCCTGATCATGCGGCTGCA 7541
QY 2472 nAlaGlyValMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeuA1 2492
DB 7542 GGCAGGCTGTCTAGGCTTCCACACCCCGGCTCCCGGAGCGGCGGCTCCG 7601
QY 2492 aGlyProHiShiA1aTTPaAPG1uG1uProTyProLeuLeuCySerGlyInTyrg1uTh 2512
DB 7602 TGGCCCCCACCACCGCTGGAGAGAGAGACCCAGCCACTGCTGTGCTGCAGTACGAGAC 7661
QY 2512 rLeuSerAPSerGlu 2517
DB 7662 ACTCTCCGACGCGAG 7677

RESULT 13

ADQ18920 ID ADQ18920 standard; DNA; 8686 BP.

XX AC ADQ18920;

XX DT 26-AUG-2004 (first entry)

DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 1739.

XX KM soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human; de.

XX OS Homo sapiens.

XX PN MO2004048938-A2.

XX PD 10-JUN-2004.

XX PF 26-NOV-2003; 2003WC-US038193.

XX PR 26-NOV-2002; 2002US-0429739P.

XX PA (PROT-) PROTEIN DESIGN LABS INC.

XX PI Aziz N, Ginsburg WM, Zlotnik A;

XX DR WPI; 2004-441208/41.

XX Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.

XX Example 2; SEQ ID NO 1739; 210pp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma
XX which comprises obtaining a first soft tissue sample from an individual
XX and a normal soft tissue sample from the same or different individual,
XX determining the expression of a gene in both samples and comparing the
XX expression of the gene in both soft tissue samples, where a higher level
XX of protein expression in the first soft tissue sample indicates the
XX presence of soft tissue sarcoma. The method of the invention has
XX cytostatic applications and may be useful for detecting soft tissue
XX sarcoma, possibly via gene therapy or vaccine production. The nucleic
XX acid sequences may be useful in diagnostic and screening applications.
XX The current sequence is that of a human soft tissue sarcoma-upregulated
XX DNA of the invention. The current sequence is not shown within the
XX specification per se but was submitted in CD format by the inventor.

XX SQ Sequence 8686 BP; 1881 A; 3069 C; 2562 G; 1174 T; 0 U; 0 Other;

Alignment Scores:

Score: 0
Percent Similarity: 12978.00
Best Local Similarity: 98.57%

Length: 8686
Matches: 2487
Conservative: 3
Mismatches: 9

Query Match: 98.21% Indels: 28
DB: 12 Gaps: 3
US-09-522-753-5 (1-2517) x ADQ18920 (1-8686)

QY 1 MesSerGlySerThrGlnLeuValAlaGlnThrTPaArgAlaThrGluProArgTyPro 20
DB 157 ATGTGGGGCTCCACACAGCCTGTGTGACACAGCGTGAAGGGCCATGAGCCCGCTACCCG 216
QY 21 ProHiSerLeuSerTyProProValGlnIleAlaArgThrHiSerThrAPValAGlyLeuLeu 40
DB 217 CCCACAGCCCTTCTCCACAGTGCAGATGCGCCGAGCGACACAGAGAGTGGGCTCTCG 276
QY 41 GluTyrglnHiShiSerArgAPTyAlaSerHiShiSerProGlySerIleIleGln 60
DB 277 GAGTACAGACACCACTCCCGGACTATGCTCCACACTGTGCGCCGCTCATATCCAG 336
QY 61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAaengluArgSerGln 80
DB 337 CCCAGCGGCGAGGCCCTCCCTGCTGTGAGTTCAGCCCGGGAATGAACGGTCCAG 396
QY 81 GluLeuHiShiLeuArgProGluSerHiSerTyLeuProGluLeuGlySerGluMet 100
DB 397 GAGCTCCACCTGGCGGCGACGATCCCACTATACCTGCCAGCTGAGGAGTCAAGATG 456
QY 101 GluPheIleGluSerTySAProArgLeuGluLeuLeuProAPProLeuLeuArgPro 120
DB 457 GAGTTCATTAAAGCAAGCCCTCGGCTAGAGCTGTGCTGACCCCTGTCTGACCG 516
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAPLeuThrTySAAPArgSer 140
DB 517 TCAACCCCTGTGGCGACGGGCGACGCTGCGGATCTGAACACCTCAACAGAGACCTGAGC 576
QY 141 LeuThrGlyLeuLeuGluProValSerProProSerProProHiSerThrAPProGluLeu 160
DB 577 CTGACGGGCAAGCTGGAAACCGGTCTCCCGCCAGCCCGGACATGACCTGAGCTG 636
QY 161 GluLeuValProProArgLeuSerTySAProGluLeuIleGlnAaMetAPArgValASP 180
DB 637 GAGCTGTGCGGCGACCGCTGTCCAGAGAGAGACTATCCAGAACATGACCGCTGTGAC 696
QY 181 ArgGluIleThrMetValGlnGlnGlnIleSerTySAProLeuTySAArgGlnGlnLeu 200
DB 697 CGAGAGATCACCATGTGAAGAGACAGATCTTAAGCTGAAGAAAGAGACAGCAACAGCTG 756
QY 201 GluGluGluAlaAlaTyProProGluProGluTyProValSerProProIleGln 220
DB 757 GAGGAGAGGCTGCCAAGCGCCGAGCTGAGAAAGCCGTGTACCGCGCCCATCTGAG 816
QY 221 SerTySAArgSerLeuValGlnIleIleTySAProGluAaArgTySAArgAlaGluAla 240
DB 817 TCGAAGCACCGGACCTGTGTGATCATTTACAGACAGAACCGGAGAGAGCTGAGAGCT 876
QY 241 AlaHiArgIleLeuGlnGlyLeuGlyProGlnValGluLeuProLeuTySAArgInPro 260
DB 877 GCAATCGGATTTCTGAAGGCTGTGGGGCCCAAGTGAAGCTGCGCTGTCAACAGCC 936
QY 261 SerAspThrArgGlnTyrglnGlnIleGlnIleGlnIleAaGlnAlaMetArgTySAArg 280
DB 937 TCGAAGACCGGAGATATGTAACATCAATAATAAACAGGCGATGCGGAGAAAGCTA 996
QY 281 IleLeuTyPheTySAArgArgAaHiSAArgTySAArgGlnTyrglnTySAArgPheCySAArg 300
DB 997 ATCTGTATCTTCAAGAGAGAGATCAAGCTCGGAAACATGAGAGCAAGAGTTCTGCCAG 1056
QY 301 ArgTySAArgGluLeuMetGluAlaLeuGluTySAArgValGluTyrglnIleGlnAaAaPro 320
DB 1057 CGCTATGACAGCTCATGAGGCTCGGAGAAAGAGTGAAGCGCATCGAAGAACACCC 1116
QY 321 ArgArgArgAlaAlaGluSerTySAArgGluTyrglnTyrglnTySAArgPheProGluIle 340
DB 1117 CGGCGGCGGCGCAAGAGAGCAAGGTTCGAGATCTTACGAGAAAGATTCCTCGAGATC 1176

QY 341 ArglySGlnArgGluLeuGlnGlnArgMetGlnSerArgValGlyGlnArgGlySerGly 360
Db 1177 CCGAAGCAGCGGAGCTGCAGAGCGCATGCGAG--AGGGTGGCCAGCGGGCAGTGGG 1233
QY 361 LeuSerMetSerAlaAlaArgSerGluHlaGluValSerGluLeuLeuLeuSer 380
Db 1234 CTGTCCATGTGCGCCCGCCGACGAGCAGAGGTGTCAAGATGTCAATGAGTGGCTCTCA 1293
QY 381 GlnGlnGlnLeuLeuGlnLeuMetArgGlnMetAlaValLeuProPheMetLeuTyr 400
Db 1294 GAGCAGGAGAACCTGGAGAAAGCAGATGCCGACGTGGCCGTGATCCCGCCAGCTGTAC 1353
QY 401 AspAlaAspGlnGlnArgIleLeuSphenLeuMetAsnGlyLeuMetAlaAspPheMet 420
Db 1354 GAGCGTGAACAGCAGCGCATCAAGTTCATCAATGAACGGGCTTATGCGCGACCCCATG 1413
QY 421 LysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGlnLysGluTrpPhe 440
Db 1414 AAGGTGTACAAAGACCGCGCATGTGAATGTGAGTGAAGCAGAGAGAGACCTTC 1473
QY 441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuLeuLeuSerPheLeuGlnArg 460
Db 1474 CCGGAGAAAGTTCATGCAGCATCCCAAGAACTTGGCTGATCGCATTCCTCGAGAGG 1533
QY 461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysAsnGluAsnTyrLys 480
Db 1534 AAGACAGTGGCTGAGTGGCTCTCTATTAATCACTGACCTGAAGAAATGAACTTAAG 1593
QY 481 SerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGln 500
Db 1594 AGCTGTGTAGAGAGAGCTATCGCGCGCGCAAGAGCCAGAGCAACAGCAGCAG 1653
QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
Db 1654 CAGCAGACAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1713
QY 521 AspGluLysGluLysGluLysGluValaGluLysGlnGluLysProGluValaGluAsn 540
Db 1714 GATGAAG 1773
QY 541 AspLysGluLysAspLeuLeuLysGluLysThrAspAspThrSerGlyLysAspAspGlu 560
Db 1774 GAAAG 1833
QY 561 LysGluAlaValAlaSerLysGlyArgGlyThrAlaAsnSerGlnGlyArgGlyLys 580
Db 1834 AAGGAGGCTGTGGCTCCAAAGCCGCAAACTGCCAAAGCCAGGAAAGCCAAAGCC 1893
QY 581 ArgIleThrArgSerMetAlaAsnGluValaAsnSerGluGluAlaIleThrProGln 600
Db 1894 CGCATCAACCGCTCAATGGCTAATGAGGCCAAGCAGAGAGAGAGAGAGAGAGAG 1953
QY 601 SerAlaGluLeuLeuLeuSerMetGluLeuAsnGluSerSerArgTrpTrpGluGluMet 620
Db 1954 AGGCGGAGCTGGCTTCATGAGCTGAATGAAGATTCTGTGAGCAAGAGAGAGATG 2013
QY 621 GlnThrAlaLysLysGlyLeuLeuGlnHlaGlyArgAsnTrpSerAlaIleAlaArgMet 640
Db 2014 GAAACAGCCAAAGAGGTCTCTGGAACAGCGCCGCACTGTGCGGCACTCCCGGATG 2073
QY 641 ValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysValArgGln 660
Db 2074 GTGGGCTCCAAAGCTGTGTGAGGTGAAGAACTTCACTTCAACTCAAGAGAGAGCAG 2133
QY 661 AsnLeuAspGluIleLeuGlnGlnHlaLysLeuLysMetGluLysGluValaGlnAlaArg 680
Db 2134 AACCTGATGATGATCTTGAGCAGCAGCAACAGTGAAGTGAAGAGAGAGAGAGAGC 2193
QY 681 ArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValVal 700
Db 2194 AGGAAAG 2253
QY 701 GluAspGluGluMetGluAlaSerGlyValSerGlyValaGlnGluMetValaGluGlu 720

Db 2254 GAGAGTGAAGAGATGAGGAGCTGCGGCGTGAAGGAGAAAGAGAGAGATGTGTGAGAG 2313
QY 721 AlaGluAlaLeuHlaSerGlyAsnGluValaProArgGlyGluCysSerGlyProAla 740
Db 2314 GCTGAH-----GCC 2322
QY 741 ThrValaAsnAsnSerSerAspThrGlnSerIleProSerProHlaSerGluAlaLys 760
Db 2323 ACTGTCAACAACAGCTCAACAGCAGACCGAGACATCCCTCTCTCACTGAGGCGGCAAG 2382
QY 761 AspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyValaAspGlyProProPro 780
Db 2383 GACACAGGAGCAGATGAGCCCAAGCCCAAGCCCAAGCTGAGCCGAGAGAGAGAGAG 2442
QY 781 GlyProProThrProPro-ArgArgThrSerArgAlaProIleGluProThrProAlaSe 800
Db 2443 GGGCCACCAACCCCAACCAACGAGAGACATCCCGCC-CCACAGATTCACCCCGGCTC 2501
QY 800 rGluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProProP 820
Db 2502 TGAAGCCACTTAAGCCCTACGCGCCCAACAGACCCCATTTCTTCACTTCCTCC 2561
QY 820 oValValProLysGluGluLysGluGluTrpAlaAlaAlaProProValaGluGlu 840
Db 2562 TGTGTCTCCCAAG 2621
QY 840 yGluGluGlnLysProProAlaAlaGluGluLeuValaValaAspThrGlyLysAlaGlu 860
Db 2622 GAG 2681
QY 860 uProValLysSerGluCysThrGluGluValaGluGluGlyProAlaLysGlyLysAspAl 880
Db 2682 GCCCTCAAG 2741
QY 880 aGluValaAlaGluValaThrAlaGluGluValaLeuLysAlaGluLysGluGluGly 900
Db 2742 GAGAGCGCTGAGAGCCAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2801
QY 900 rGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerAlaTh 920
Db 2802 CGCAGAGGCCACCAAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2861
QY 920 rCysSerAlaAspGluValaAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerP 940
Db 2862 CTGCACTGCAAGAGAGTGAATGAGCGCGAGCGCGCGAGCAGAAACCGGCTGTCTCCC 2921
QY 940 aArgProSerLeuLeuThrProThrGlyAspProAlaAlaAsnAlaSerProGlnLysP 960
Db 2922 AAGGCCAGCTCTCAACCCGAGCTGAGCAGACCCCGAGCCAAATGCTCACCCAGAGCC 2981
QY 960 oleuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLy 980
Db 2982 ACTGAGCTGAAGACACTGAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3041
QY 980 sValHlaGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaProP 1000
Db 3042 AGTCCATGAGCCCGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3101
QY 1000 oProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlnLysSerProAr 1020
Db 3102 GCCACCGCAAAACCTCAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3161
QY 1020 gGlyLysSerArgSerProAlaProProAlaAspLysGlu----- 1033
Db 3162 GGGCAAGAGCAG 3221
QY 1034 ---AlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerG 1052
Db 3222 CCGAGCTTGTGAGCGAG 3281
QY 1052 yLeuProPheProValProProArgGluValaIleLysAlaSerProHlaProAspPr 1072

Dh 3382 CTTGCCCCCTGCGCCCCCGGTGAGTGTATCAAGGCTTCCCGCATGCCCCGAGACC 3341
Qy 1072 oSeRAlApheserTyrAlaProProGlyYHiaProleuProleuGlyLeuHiaaPThraI 1092
Db 3342 CTCAGCCTTCTCTTAAGCTTCAACCTTGACCACTGCCCCCTGGCTCTCAAGACAGTGC 3401
Qy 1092 aArgProValLeuProArgProProThriIseSerAsnProProleuIseSerAl 1112
Db 3402 CCGGCGCGTCTGCGCGGCCACCAACCATCTCCAAACCGGCTCCCTCATCTCTGCG 3461
Qy 1112 alyvHiAProSerValLeuGlyuaArgInIleGlyAlaIleSerGInGlyMeSerValG 1132
Db 3462 CAACACCCCCAGCCTCTCTGAGAGCAAAATAGTGCCATCTCCAAAGAAAGTGGTCCA 3521
Qy 1132 nLeuHiaValProTyrSerGInuHiaAlaValaProValGlyProValThrMetGlyLe 1152
Db 3522 GCTTCAGTCTCCGTAAGCTCAAGCATGCGAAGCCCCGGTGGCCCTGTGACACATGGGCT 3581
Qy 1152 uProleuProMetAsnProLySlyLeuAlaProPheserGlyValIlySGInGInLe 1172
Db 3582 GCCCTGCCCCATGAGCCCAAAAAGCTGGCACTTCAAGCGAGTGAAGCAAGAGCT 3641
Qy 1172 uSerProArgGlyGInaIagIyProProGlyuSerLeuGlyValProThraIaGInGuaI 1192
Db 3642 GTCCCAAGGGGCGAGGCTGGGCCACCGAAGGCTGGGGGTGCCCAAGCCCAAGAGGC 3701
Qy 1192 aSerValLeuArgGlyYHriaIleuGlySerValProGlyGlySerIleThrySlyGlyI 1212
Db 3702 GTCCGTCTGAGAGGAGCAAGCTTGAGCTCAAGTTCGGGCGAAGCATCAACAAAGCAT 3761
Qy 1212 eProSerThraArgValProSerAspSerAlaIleThryArgGlySerIleThriHiaG 1232
Db 3762 TCCAGACACAGGGTGCCCTCGACAGCGCATCATACCGCGGCTCTCAACCCACG 3821
Qy 1232 YThrProAlaAspValLeuTyrIlySGlyYHriaIleThraArgIleIleGlyIleuAspSerPr 1252
Db 3822 CACCCAGCTGACCTCTCTGTAAGGACACCATCAACAGATCTCGGCGAGACACGCC 3881
Qy 1252 oSeRArgLeuAspArgGlyYHriaIleuAspSerLeuProLySlyIleValIleTyrGInG 1272
Db 3882 GAGTGGCTTGAGACCGGCGCGGAGACAGCTGCCCAAGGGCCACGTCATCAAGAAAG 3941
Qy 1272 YLySlySGlyYHriaIleuSerTyrGInGlyGlyMeSerValThriGlySerIlySG 1292
Db 3942 CAAGAAAGGCGCAAGCTTGTCTTATGAGGGTGGAGTGTCTGTACCAAGTCTCAAG 4001
Qy 1292 uAspGlyArgSerSerGlyProProThriIleGlyThraAlaProLyAspArgThryrAs 1312
Db 4002 GGAACGAGAAAGCAAGCTCAGGACCCCCCATGAGACGGCGCCCAAGCGCACCTATGA 4061
Qy 1312 pMetMetGlyGlyYHriaIleuArgAlaIleSerSerAlaSerIleGInGlyLeuMetG 1332
Db 4062 CATGATGAGAGGCGCGCTGGGAGAGCATCTCTCAAGCCAGCATCAAGGTCTCAAGGG 4121
Qy 1332 YATGAlaIleProProGlyIleuArgHiaSerProHiaIleuLySlyGInHiaIleAr 1352
Db 4122 CCGTGCCATCCCGCGGAGGACACAGCCCCACCATCTCAAGAGAGACACCACTCG 4181
Qy 1352 gGlySerIleThriGInGlyIleProArgSerTyrValGInaIagInIleuAspTyrLeuAr 1372
Db 4182 CCGGTCCATCAACAAAGGATCCCTCGGTCTTACGTGAGGACAGAGAGCATCTCGCG 4241
Qy 1372 gATGAGlyAlaIleuLeuLeuLyAspArgGlyYHriaIleuProProProProProSerArgAs 1392
Db 4242 TCGGAGGCGCAAGCTCTTAAAGCGGAGGAGCAAGCTCCGCGCCACCGGCTCAAGGGA 4301
Qy 1392 pLeuThriGlyAlaIleTyrIleThriGInaIleuGlyProLeuLyLeuLyAspProAlaHiaSG 1412
Db 4302 CTGAGCGGAGGCTCAAGAGAGAGCGGCTGGGCCCCCTGAAGGTGAAGCGGCGCATGA 4361
Qy 1412 uGlyLeuValAlaThraValIlySGlyAlaGlyArgSerIleHiaGInIleProArgGInG 1432
Db 4362 GGGGCTGTGGTCCAGGCTGAAAGAGGCGGGCGGCTCATCATGATGATCCGCGCGAGGA 4421

Qy 1432 uLeuArgHiaThraProGlyLeuProleuAlaProArgProleuLySGlySerIleTh 1452
Db 4422 GCTGCGGACACGCGCCAGAGCTGCCCCCTGGCCCCCGGCGCTCAAGAGGCTCATATAC 4481
Qy 1452 rGInGlyThraProleuLySlyrAspThriGlyAlaIleSerThriGlySerIlySlyHiaS 1472
Db 4482 GCAAGGACACCGGCTCAAGTGAACACCGGCGCTCCACCACTGGCTCCAAAGACAGA 4541
Qy 1472 pValArgSerIleuIleGlySerProGlyYHriaGThraPheProProValHiaProleuAspVa 1492
Db 4542 GTAGCTCTCTCATGCGAGAGCCCGGCGGACCTTCCACCGCTGACCCGCTGGAGT 4601
Qy 1492 lMetAlaAspAlaArgAlaLeuGlyuaArgAlaCyTeTyrGInGlyuSerLeuLySserArgPr 1512
Db 4602 GATGGCGGACCGCGGCGACTGGAACCTGCTCTTACAGAGAGAGCTGAAGAGCGGCC 4661
Qy 1512 oGlyThraIleSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProG 1532
Db 4662 AGGAGCGCGAGAGCTCGGGGGGCTCATTTGGCGCGCGGCTGATTTGGCTGA 4721
Qy 1532 uLeuGlyLyProArgInSerProleuThryrGlyIleuAspHiaGlyAlaProPheAlaG 1552
Db 4722 GCTGGTAAAGCGGCGAGAGCCCTGACCTTGAAGAGACACAGGCGACACCTTTGGCGG 4781
Qy 1552 YHriaLeuProArgGlySerProValThrMetArgGlyuProThraProArgLeuGInG 1572
Db 4782 CCACCTCCACAGAGTTCGCCCTGACCATGCGGAGGCCACCGCGGCTTCAAGAGG 4841
Qy 1572 ySerLeuSerSerSerIlyAlaSerGInaAspArgLyLeuThraSerThraProArgGIn 1592
Db 4842 CAGCTTTCGTTCAGAGCAAGGATCCCAAGACCAAGACTGACGTGAGCGCTGTGAT 4901
Qy 1592 eAlaLySserProHiaSerThraValProGlyIleHiaIleProHiaProIleSerProTyrG 1612
Db 4902 CGCAAGTCCCGGACAGCACCGTGCGGAGCACACCCACCTCTGCGCTTATGA 4961
Qy 1612 uHiaIleuLeuArgGlyYHriaIleuValaAspLeuTyrArgSerHiaIleProleuAlaPh 1632
Db 4962 GCACCTCTTCGGGGCGTGAAGCTGAGCTGTATCGAGCCACATCCCCCTGGGCTT 5011
Qy 1632 eAspProThraSerIleProArgGlyIleProleuAspAlaAlaIleTyrTyrLeuPr 1652
Db 5022 GACCCCATCTCATATACCCCGGCGATCCCTGTGAAGAGCGGCTGCTATCACTGCC 5081
Qy 1652 oArgHiaLeuAlaProAsnProThryrProHiaIleuTyrProProTyrTyrLeuIleArg 1672
Db 5082 CCGACACTGGCCCCCAACCCCACTAACCGGACCTGTATCCACCTTACCTCATCCGCG 5141
Qy 1672 YTyProAspThraAlaIleuGlyIleuAspArgGInThriIleIleAsnAspTyrIleThrSe 1692
Db 5142 CTACCCCGACAGGCGGCGCTGAGAAACGGAGAACCATATCATATCATATCATCTTC 5201
Qy 1692 rGInGInMetHiaIleAsnThraIleThraIleAlaGInaArgAlaAspMetLeuArgG 1712
Db 5202 GAGAGAGATGACACCAACAGGCGCAACCGGCAATGGCCCAAGAGCTGATGTGAGAGGG 5261
Qy 1712 YLeuSerProArgGlyuSerSerIleuAlaLeuAsnTyrAlaIleGlyProArgGlyIleI 1732
Db 5262 CCTCTGCCCCCGGAGAGCTCTGCTGGCACTCACTACGTCGGGGTCCCGAGGACATCAT 5321
Qy 1732 eAspLeuSerGInValProHiaIleuProValaIleuValaProProThraProGlyThraProAl 1752
Db 5322 GAGCTGTCCCAAGTGCACACTCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5381
Qy 1752 aThraIleMetAspArgIleuAlaTyrLeuProThraIaProGInProPheserSerArgH 1772
Db 5382 CACCGCATGAGACGCGCTTGTACCTCCCAACCGGCGCCCAAGCGCTTCAAGAGCGGCA 5441
Qy 1772 aSerSerSerProleuSerProGlyGlyProThriHiaIleuThryLyProThriThryrSe 1792
Db 5442 CAGCAGCTCCCACTCTCCCAAGAGGTCCAAACACTTGAACAAACCAACCAACAGTGC 5501

QY 1792 rSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluArg 1812
DB 5502 CTGCTCCGAGCCGGAGCGAGACCGGGATGTGAGAGCGGGATCGGGAGCGGGAAAA 5561
QY 1812 sSerTlleuThrSerThrThrValGluHisAlaProL]eTPArgProGlyThrG 1832
DB 5562 GTCCATCCTCAGCTCAGACGACGAGGTGAGACGACCCATCTGAGACCTGTGTCAAA 5621
QY 1832 uGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProL 1852
DB 5622 GAGAGACGAGCGAGACGAGCGAGCGCGGGGTGGGGGACGACGAGCGCGCCCG 5681
QY 1852 aSerHisSerHisAlaHisGlnHisSerProL]eSerProArgThrGlnAspAlaLeuG 1872
DB 5682 CTCCCACTCCCATGCCACACGACACTCGCCATCTCCCTCGGACCCAGGATCCCTCCA 5741
QY 1872 nGlnArgProSerValLeuHisAlaThrGlyMetGlyGlyLeIleThrAlaValGluP 1892
DB 5742 GAGAGAACCCAGTGTCTTCAACAACAGGATGAAGGATATCATCACCGCTGTGAGCC 5801
QY 1892 oSerLysProThrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaTh 1912
DB 5802 CAGCAAGCCCAAGCTCTGAGGTCCACCTTCACTCTCACTCCCTTCCGCACTGCGCAC 5861
QY 1912 rPheProProAlaThrHisCysBProLeuGlyGlyThrLeuAspGlyValTyBProThle 1932
DB 5862 ATTCCCACTGGACCCACCTGAGCCCACTGGGCGGACCTCGATGGGGTACCTTACCTT 5921
QY 1932 uMetGluProValLeuLeuProLysGluAlaProAlaGValAlaArgProGluArgProAr 1952
DB 5922 CATGGAGCCGCTTCTCTGCGCAAGAGGCGCCCGGCTGCGCGGCGAGAGGCGCGCG 5981
QY 1952 gAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAl 1972
DB 5982 AGCAGACACCGGCGCAAGCTTCTTCTCGCAAGCCCGGACCGCGCTCGGGCTGAGCGCGC 6041
QY 1972 aSerSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAl 1992
DB 6042 CTCTCTCCCGAGCAAGGGCTCGAGCCCGGCGCTTGTCTCTGTCTGTGCGCACGC 6101
QY 1992 aThrTlleAlaArgThrProAlaLysAlaLeuAlaProHisAlaAlaSerProAspProP 2012
DB 6102 CACCATCGCGCGACCCCTGCGAAGAACTCGCACCTCACACGCGGACCGGACCGCGC 6161
QY 2012 oAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPhe 2032
DB 6162 GCGGCACTGCTGCTCGGCTCGGACCGGACCGGAGAAAGACTCAAAAGTAAACCTTTTC 6221
QY 2032 rLleGlnGluLeuGluLeuArgSerLeuGlyTyHisGlySerSerTySerProGluG 2052
DB 6222 CATCCAGGAAGTGGAACTCGCTTCTGTGGGTACACAGGCAAGCTACAGCCCGGAGG 6281
QY 2052 yValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLy 2072
DB 6282 GGTGGAGCCCGTCAAGCTGTGACTCACCACTGTGACCCACAGCAAGGGGCTCCCAA 6341
QY 2072 sHisLeuGluGluLeuAspLysSerHisLeuGluGluGluLeuArgProLysGlnProG 2092
DB 6342 GACCTGGAAGAACTGACAGAGCAAGCACTGTGAGGGGAACTGCGGCCCAAGAGCCAGG 6401
QY 2092 yProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProLys 2112
DB 6402 CCCCGTAGAGCTTGGCGGGAGGCGCGCCACCTCCCACTCGGGCGCTGCTGAGAG 6461
QY 2112 rGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgVa 2132
DB 6462 CAGGCCCTCGTCAAGCCCGCTGTCCAGAGCGCGCCAGGGGTCAAAAGTACCAAGCGGCT 6521
QY 2132 lValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyThrArgHisIsp 2152
DB 6522 GGTCAACCTTGCGGACGACATCGTAGGTCAATCAACAGAACTACACCGGACCAACCC 6581
QY 2152 oGlnGlnLeuSerAlaProLeuProAlaProLeuTySerPheProGlyAlaSerCysPr 2172

DB 6582 ACAGCAGCTCACGCGACCCCTGCGCGCGCGCTTACTCTTCCCTGGGCGCAGCTGCC 6641
QY 2172 oValLeuAspLeuArgArgProProSerAspLeuTyTleuProProProAlaHisAl 2192
DB 6642 CGTCTGAGACTCCGCGCGCGCCACCAAGTACTTACTCTCCGCGCGCGGACCATGTGC 6701
QY 2192 aProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnLysTh 2212
DB 6702 CCGGCGCGGTGCTCCCGCCACAGCGAAGGGGCAAGAGTCTCCAGAGCCAAACAAGAC 6761
QY 2212 rSerValLeuGlyGlyGlyGluAspGlyLleGluProValSerProProGluGlyMetTh 2232
DB 6762 GTCGGTCTGGGTGGGTGAGAGCCGTAATTGAACCTGTGTCCCGCCAGAGGCAATGAC 6821
QY 2232 rGluProGlyHisSerArgSerAlaValTyProLeuLeuTyArgAspGlyGlyGlnTh 2252
DB 6822 GAGGCCAGGACACTCCCGAGTGTGTGTACCGCTGTGTACCGGAGTGGGACAGAC 6881
QY 2252 rGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPheP 2272
DB 6882 GAGGCCACAGGATGGCTCCAAAGTCTCACGAGCAACACAGCCAGCCGCAAGCTTCTT 6941
QY 2272 eSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluLleAsnLy 2292
DB 6942 CAGCAAGCTGACCGAGAGCACTCCGCGCATGTGCAAGTCAAGAAACAAAGATCAACA 7001
QY 2292 sLysLeuAsnThrHisAsnArgAsnGluProGluTyRAsnLleSerGlnProGlyThrG 2312
DB 7002 GAACTGAAACCCCAACACCGAATGAGCTGAATCAATATACGACCAAGCTGGACGGA 7061
QY 2312 uLlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyArgSerGlnAlaVa 2332
DB 7062 GATCTTCAATATAGCCCGCATCATCCGGAACAGGCTTTATGACTATAGAAGCAGCGGCT 7121
QY 2332 lGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetG 2352
DB 7122 GAGAGAACATGCGACGACCAACATGGGGCTGAGGCGCAATTAAGAAGCACTCATGG 7181
QY 2352 yLysTyRAspGlnTyRArgGluGlySerProProLeuSerAlaAsnAlaPheAsnProLeuAs 2372
DB 7182 TTAATATGACCAAGTGGAAAGATCCCGCGCTGACGCGCAATGCTTTTAACCTTGAA 7241
QY 2372 nAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHi 2392
DB 7242 TGCAGTGCAGGCTGCGCGCTGTATGCCAATAACCGCTGTGAGCGAGAGTACCA 7301
QY 2392 sThrLeuThrSerProGlyGlyGlyGlyLysValLysValSerGlyArgProSerSer 2412
DB 7302 CACACTCACCTGCGCAGGTGGCGGCGGAAAGGCMAAGTCTGTGCGACACCCAGCGCGG 7361
QY 2412 gLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSe 2432
DB 7362 AAAAGCAAGTCCCGCGCGCGCGCTGGATCTGGAGAACCGGCAACCTCTGTCTCTC 7421
QY 2432 rValHisSerGluGlyAspCysAsnArgTrgThrProLeuThrAsnArgValTrpGluAs 2452
DB 7422 AGTGCACTGGAGGAGATGCAACCGCGGACGCGCTCACCAACCGCGTGGAGAGA 7481
QY 2452 pArgProSerSerAlaGlySerThrProPheProTyTyAsnProLeuLleMetArgLeuG 2472
DB 7482 CAGGCCCTGCTCGCGAGTTCACAGCAATTCCTTCAAAACCCCTATATCATGGGCTGCA 7541
QY 2472 nAlaGlyValMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeuAl 2492
DB 7542 GCGGGGTGTATGTCTTCCCAACCCCAACCGGCGCTCCCGCGGAGACGGGCGCTCTGC 7601
QY 2492 aGlyProHisAlaAlaTrpAspGluGluProLysProLeuLeuCysSerGlnTyRArgL 2512
DB 7602 TGGCCCCCAACAGCTGGGAGCAGAGAGCCAAAGCCACTGTCTGTCTGCAATACGAGAC 7661
QY 2512 rLeuSerAspSerGlu 2517

DB 7662 ACTCTCCGACGACGAG 7677

RESULT 14
ADL12577
ID ADL12577 standard, cDNA; 9053 BP.
XX
XX ADL12577;
AC
XX
XX 06-MAY-2004 (first entry)
DT
XX
XX Human steroid-induced C3A liver cell. cDNA #306.
DE
XX
XX ss; gene; Hepatotropic; Gene therapy; Wilson disease; liver disorder;
KM steroid therapy; cirrhosis; hepatitis; human; C3A liver cell.
XX
XX Homo sapiens.
OS
XX
XX US6673549-B1.
PN
XX
XX 06-JAN-2004.
PD
XX
XX 12-OCT-2001; 2001US-00976594.
PF
XX
XX 12-OCT-2000; 2000US-0240409P.
PR
XX
XX (INCYTE CORP.
PA
XX
XX Furness LM, Buchbinder JL;
PI
XX
XX MPI; 2004-068610/07.
DR
XX
XX
XX Combination useful for preparing a composition for treating liver
PT disorders associated with steroid therapy, e.g., cirrhosis or hepatitis;
PT comprises cDNAs that are differentially expressed in response to steroid
PT treatment.
XX
XX Claim 1; SEQ ID NO 306; 141pp; English.

The invention relates to a combination comprising cDNAs that are
CC differentially expressed in response to steroid treatment. Also included
CC are the following: a high throughput method for using a cDNA to detect
CC differential expression of nucleic acids in a sample; and a high
CC throughput method of screening molecules or compounds to identify a
CC ligand that specifically binds a cDNA. The sample is from a subject with
CC Wilson disease and comparison of a standard defines a stage of that
CC disease. The high throughput method of screening molecules or compounds
CC to identify a ligand that specifically binds a cDNA comprises: combining
CC the combination with molecules or compounds under conditions to allow
CC specific binding; and detecting specific binding between each cDNA and at
CC least one molecule or compound. The molecules or compounds are regulatory
CC proteins. The combination is useful for preparing a composition for
CC treating liver disorders associated with steroid therapy, e.g., cirrhosis
CC or hepatitis. The present sequence represents a human cDNA which is
CC differentially expressed in steroid-induced C3A liver cells. Note: The
CC sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html.

XX
XX Sequence 9053 BP; 1904 A; 3171 C; 2722 G; 1254 T; 0 U; 2 Other;
SQ

Alignment Scores:

Pred. No.:	0	Length:	9053
Score:	12922.00	Matches:	2487
Percent Similarity:	98.26%	Conservative:	2
Best Local Similarity:	98.18%	Mismatches:	10
Query Match:	97.78%	Indels:	35
DB:	12	Gaps:	5

US-09-522-753-5 (1-2517) x ADL12577 (1-9053)

QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTyrArgAlaThrGluProArgTyrPro 20
DB 520 ATGTCCGGATCCACACAGCCTGTGGCAAGACGTGAGGGCCACCTAGGCCCGCTACCCG 579

QY 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisArgPheValGlyLeuLeu 40
DB 580 CCCACAGCCTTTCTTACCAGTGCAGATGCCCCGAGCAGACAGCGTGGGCTCTCTG 639

QY 41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
DB 640 GAGTACCAGACACACTCCCGGACTATGCTCCACCTGTGCGCCGCTCATCATTCAG 639

QY 61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyPheGlnAlaGlySerGln 80
DB 700 CCCACGCGGAGGCGCTCCCTGCTGTGAGTTCCAGCCCGGAGATGAACGATCCAG 759

QY 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLeuSerGluMet 100
DB 760 GAGCTCACCTGGCGGCAAGTCCCATCATCTCTCCGAGCTGGGAGAGTGAAGATG 819

QY 101 GluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
DB 820 GAGTTGATTGAAGAAGAGCGCTCGGCTAGAGCTGCTGCTGACCCCTGTGCGACCG 879

QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
DB 880 TCACCCCTGCTGCGCACGCGGCCAGCTTGGCGGATCTGAAGACCTCACAAAGACCGTACG 939

QY 141 LeuThrGlyLysLeuGluProValSerProProSerProProHisArgAspProGluLeu 160
DB 940 CTGACGGGCAAGCTGGACCGGTGTCTCCCGCCAGCCCGCCGACACTGACCTTGAGCTG 999

QY 161 GluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAsp 180
DB 1000 GAGCTGTGCGCGCACGCGCTGTCCAAAGAGAGCTGATCAGAAACATGGACCGCGTGAC 1059

QY 181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeu 200
DB 1060 CGAGAGATCACCATGTGTAGACGACGAGATCTTAACTGAAAGAAAGACGAAACGCTG 1119

QY 201 GluGluGluAlaAlaLysProProGluProGluLysProValSerProProIleGlu 220
DB 1120 GAGGAGAGAGCTGCCAAGCGCCCGGAGCTGAGAGCCGTGTACCGCCGCTCATCGAG 1179

QY 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysValGlnAla 240
DB 1180 TCGAAGCACCGGACGCTGTGAGATCATCTACGAGAACCGGAAGAGGTGAAGCT 1239

QY 241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrArgGlnPro 260
DB 1240 GCACATCGGATTTCTGAAAGGCTTGGGGCCCGAGGTGAGCTGCCGTGTACACCGAGCC 1299

QY 261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
DB 1300 TCGACACCGGAGATGATATGAAACATCAAAATTAACAGAGCTGGAGAAAGCTA 1359

QY 281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGln 300
DB 1360 ATCTTGTACTTCAAGAGAGGAATCACGCTCGGAACAATGGAGAGCAGAAGTTCTGCCAG 1419

QY 301 ArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnPro 320
DB 1420 CGCTATGACCACTCATGAGAGGCTTGGGAGAAAGAGTGAAGCGCATCAAGAACACCCC 1479

QY 321 ArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340
DB 1480 CCGCGGCGGCGCAAGAGAGCAAGGTGCCGAGTACTACGAAAGCAAGTTCTCTGAGATC 1539

QY 341 ArgLysGlnArgLysGlnGluArgMetGlnSerArgValGlyGlnArgLysSerGly 360
DB 1540 CCGAAGCAGCGGAGCTGCAAGAGCGCATGACAGAGCGGTGGGCAAGCGGCAAGTGGG 1599

QY 361 LeuSerMetSerSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
DB 1600 CTGTCTATGTGGCGCGCCGCGAGCGAGCAGCGAGGTGTCAAGATCATTCATGCGCTCTCA 1659

```
QY 361 GtugInguAaenLeugLulysGImeArvgInleuAaIaValIleProProMetLeuTyR 400
Db 1660 GAGCGAGGAAACCGGAAAGACAGATGGCCGACGTGGCGGTGATCCCGCCATGCTGTAAC 1719
QY 401 AAspAaAspGInguAaIlelysbheIleAenMeAenGlyLeuMeAlaAspProMet 420
Db 1720 GACGGCTGACACAGCGGATCAAGTTTCATCAACATGAAGAGGACTTATGGCCCAATG 1779
QY 421 LysValTyTyAspAspArgInValMeAenMeCTPSeRgInguInguLulysGluThrPhe 440
Db 1780 AAGGTGTAACAAGACCGCCAGGTCAATGAACATGTGAGTGAAGACAGAGAAAGACCTTC 1839
QY 441 ArgGluTyPheMeGImhiAspProLysAenPheGlyLeuIleAlaSerPheLeuGluArg 460
Db 1840 CGGGGAAGGTTCAATCAGATCCCAAGAACTTTGGCTGATCGCATTCATTCCTGAGAGG 1899
QY 461 LysThrValAlaGluCyValLeuTyTyTyTyLeuThrTyLeuAspAspGluAsnTyPlys 480
Db 1900 AAGACAGTGGCTGAATGGCTCTCTATTATTACTGACTAAGAAAGATGAGAACTATAG 1959
QY 481 SerLeuValArgArgSerTyArgArgArgGlyTySerGInguIn---GInguInguIn 499
Db 1960 AGCCGTGTGAGACGAGAGCTATGGGGCCGCGCAAGACGACAGCAACGACAGCAG 2019
QY 500 GInguInguInguInguInguInguInguInProMetProArgSerSerGInguIngu 519
Db 2020 CAGCGACGACAGCAGCAGCAGCAGCAGCAGCAGCCCATGCCCCGAGCAGCAGAGAGG 2079
QY 520 LysAspGluTyGluTyGluTyGluTyGluTyGluTyGluTyGluTyGluTyGluTyGlu 539
Db 2080 AAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2139
QY 540 AsnAspTyGluAspLeuLeuTyGluTyTyThrAspAspThrSerGlyGluAspAsnAsp 559
Db 2140 AACGACAAAGAAAGACCTCTCAAGAGAGAAACAGACGACCTCAGGGAGAGACAAACAG 2199
QY 560 GtulyGluValaValAlaSerTyGlyTyArgTyThrAlaAsnSerGInguTyArgTyGly 579
Db 2200 GAGAAAGAGGCTGTGCTCCCAAGAGCCGCAAAACTGCGCAACAGCAGGAGAGAGAGCAAA 2259
QY 580 GlyArgGlyeThrArgSerMetAlaAsnGluValaAsnSerGluGluValaIleThrProIn 599
Db 2260 GCGCCCATCAACCCGCTCAATGACTAATGAGGCCAACAGGAGAGAGCCATCACCCCCAG 2319
QY 600 GInserAlaGluLeuAlaSerMetGluLeuAenGluSerSerArgTyThrGluGluGlu 619
Db 2320 CAGACGCGCGAGCTGGCTCCATGAGCTGAATGAGAGTTCTCGCTGACAAAGAGAA 2379
QY 620 MetGluThrAlaTyLeuTyGlyLeuLeuGluValGlyTyArgAsnTrpSerAlaIleAlaArg 639
Db 2380 ATGGAAACAGCCCAAAAGGTCTCCTGGAAACAAGCCGCGCACTGGTCCGCCATCCCGG 2439
QY 640 MetValGlySerTyThrValSerGInCyTyLeuAsnPheTyPheAsnTyTyTyPlysArg 659
Db 2440 ATGTGTGGGCTCCAAAGCTGTGTGCGAGTGTAAAGACTTACTTCACTACAAAGAGAG 2499
QY 660 GInaenLeuAspGluIleLeuGInguInhiTyLeuTyGlyMeGluTyGluTyArgAsnAla 679
Db 2500 CAGAACCTCGATGAGATCTTGACAGACGACAAAGCTGAAGATGAGAGAGAGAGAAACGCG 2559
QY 680 ArgTyGlyTyLeuTyLeuTyAlaProAlaAlaAla-SerGluGluValaAlaPheProProVa 699
Db 2560 CGGAGAAAGAAAGAAAGAGCGCGCGCGCGCGCGCGAGAGAGGCTGCAATTCGCCCGCT 2619
QY 699 IValGluAspGluGluMetGluAlaSerGlyValaSerGlyAspAspGluGluMetValGlu 719
Db 2620 GGTGAGGATGAGAGATGAGAGCGCTCGGCGCTGACCGGAAATGAGAGAGAGATGGTGA 2679
QY 719 uGluTyGluValaLeuValaSerGlyAsnGluValaProArgGlyGluTySerGlyPr 739
Db 2680 GAGAGCTGAA----- 2689
QY 739 cAlaThrValaAsnAsnSerSerAspThrGluSerIleProSerProhiIeThrGluAlaAl 759
Db 2690 -GCCACTGTCAACAACACACTAGACACAGAGAGATCCCTCTCTCACTGACAGGCGCGC 2748
QY 759 AlysAspThrGluGInaenGlyProLysProProAlaThreInguValaAsp-GlyProP 779
Db 2749 CAAGAGACACAGGACAGATGGGCCCAAGCCCCCAGCCACCTGTGGCGCGAGAGGGCCAC 2808
QY 779 rOProGlyProProThrProPro-ArgArgThrSerArgAlaProIleGluProThrPro 798
Db 2809 CCCCAGGCGCCACCCACCCACACCGAGAGACATCCCGGCC-CCCACTGAGGCCACCCCG 2867
QY 799 AlaSerGluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaPro 818
Db 2868 GCCTTGAGGCCACCGAGAGCCCTTAGCCCCCAAGACAGACCCCAATCGCCTCTGACCT 2927
QY 819 ProProValaProLysGluGluTyGluTyGluTyGluTyGluTyGluTyGluTyGluTyGlu 838
Db 2928 CCTCTGTGTGTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2987
QY 839 GtugTyGluGluGluTyProProAlaAlaGluGluLeuAlaValAspThrGlyTyVala 858
Db 2988 GAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3047
QY 859 GtugTyProProValaLysSerGluTyThrGluGluGluGluGluGluGluGluGluGlu 878
Db 3048 GAGAGAGCCGCTCAAGAGCGAGTGCACGAGAGAGCGAGAGAGAGAGAGAGAGAGAGAG 3107
QY 879 AAspAlaGluAlaAlaGluAlaThrAlaGluGluTyAlaLeuTyAlaGluTyGluTyGlu 898
Db 3108 GACGGGAGAGCCGCTGAGAGCCAGCGCGAGGGGGGCTCAAGAGCAGAGAGAGAGAGAG 3167
QY 899 GlySerGlyTyArgAlaThrThrAlaLysSerGlyAlaProGluAspSerAspSerSer 918
Db 3168 GAGAGAGCGCAGAGGCCACCAAGAGCTCGGGCGCCCCCAAGAGAGAGAGAGAGAGAG 3227
QY 919 AlaThrCySerSerAlaAspGluValaAspGluValaGluGluTyGlyAspTyAsnArgLeuLeu 938
Db 3228 GCTACTGACAGTGCAGAGAGAGTGAAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3287
QY 939 SerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlu 958
Db 3288 TCCCCAAGGCCACGCTCTCTCAACCCCGAGCTGGCGAGCCCCCGGCGCAATGCTCACCCAG 3347
QY 959 LysProLeuAspLeuTyGluTyGluTyGluTyGluTyGluTyGluTyGluTyGluTyGlu 978
Db 3348 AAGCCACTGAGACTGAAACAGCTGAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3404
QY 979 ThrTyValaIleGluProProArgGluAspAlaAlaPheProThrTyProAlaProPro-Al 998
Db 3405 ACCAAAGTCCATGAGCCCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3464
QY 998 AProProProProGluInaenLeuGluInProGluSerAspAlaProGluInProGlySerSe 1018
Db 3465 CCGACCGGCAACCGCAAAACCTGACAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3524
QY 1018 rProArgGlyTySerArgSerProAlaProProAlaAspTyGlu----- 1033
Db 3525 CCCCCGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3584
QY 1034 -----AlaPheAlaAlaGluAlaGluTyLeuTyLeuTyLeuTyLeuTyLeuTyLeuTy 1050
Db 3585 GTTCTTCCAGCTTCGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3644
QY 1050 rSerGlyLeuProPheProProAlaProProArgGluValaIleTyValaSerProhiAlaAla 1070
Db 3645 TTCCGGGCTGGCTTCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3704
QY 1070 cAspProSerAlaPheSerTyAlaProProGlyValaProLeuProLeuTyGluTyGluTyGlu 1090
Db 3705 GAGCCCTCAAGAGCTTCTCTAAGCTGACCTGATCAACACAGAGAGAGAGAGAGAGAGAGAG 3764
QY 1090 pThrAlaArgProValaLeuProArgProProThrIleSerAsnProProProLeuIleSe 1110
```

Db 3765 CACTGCCGGCCCGTCTGCCGGCCACCACCATCTCCAAACCCGCTCCCTCATCTC 3824
Qy 1110 rSerAlaLysHisProSerValLeuGluAArgGlnIleGlyAlaIleSerGlnIleMetSe 1130
Db 3825 CTCTGCCAAGACACCCCGAGCTCTCCAGAGGCAATAGTGCCATCTCTCCAGGAATGTC 3884
Qy 1130 rValGlnLeuHisValProTyrSerGlnHisValAlaValProValGlyProValThrMe 1150
Db 3885 GGTCCAGCTCCAGTCCCGTACTCAGAGCATGCCAAGGCCCGGTGGGCTGTCAACAT 3944
Qy 1150 rGlyLeuProLeuProMetAspProLysIleValAlaProPheserGlyValIleGlnI 1170
Db 3945 GGGGCTGCTCCCTCCATGGACCCCAAAAGCTGGCACCTTCCGCGAATGAAAGCGGA 4004
Qy 1170 uGlnLeuSerProArgGlyGlnIleGlyProProGluSerLeuGlyValProThrAlaG 1190
Db 4005 GCAGCTGTCCCAAGGGGCCAGGCTGGGCCACCGAAGCTTGGGGTGGCCACAGCCCA 4064
Qy 1190 ngIuaIAserValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrI 1210
Db 4065 GAGAGGCTCGGTGCTGAGAGGACAGCTGGGCTCATGTTCCGGGCGAAGCATCACAA 4124
Qy 1210 rGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleTh 1230
Db 4125 AGGCATTTCCAGACACGGGTGCTCTCGACAGCGCATCATACCGCGCTTCATCAC 4184
Qy 1230 rHisGlyThrProAlaAspValLeuTyrIleValGlyThrIleThrArgIleIleGlyI 1250
Db 4185 CCAGGCAAGCCAGCTGACCTCTGTCAGAGGACCATACACAGGATCATCGGCGAGA 4244
Qy 1250 pSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyr 1270
Db 4245 CAGCCCGAGTCGTTGGACCGCGCGGAGACAGCTCCCAAGGCCACATCATCTA 4304
Qy 1270 rGlyGlyIleValGlyHisValLeuSerTyrGlnGlyIleMetSerValThrGlnCysSe 1290
Db 4305 CGAAGGCAAGAAAGGCGCATCTTGTCTCTATGAGGGTGCGCATGTCTGTGACCCAGTGCTC 4364
Qy 1290 rLysGluAspArgLysSerSerSerGlyProProHisGluThrAlaAlaProLysArgTh 1310
Db 4365 CAAAGAGACGGAGAAAGCTCTAGAGACCCCTCATGAGACGCGCCCAAGGCGCAC 4424
Qy 1310 rTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLe 1330
Db 4425 CATATGACATGATGAGAGGCGCGGTGGAGGCACTCTCTCAAGCAGCATCGAAGTCT 4484
Qy 1330 uMetGlyArgAlaIleProProGluLysArgHisSerProHisIleLeuLysGlnIleHis 1350
Db 4485 CATGGCCCGTGCATCCCGCGAGGACACACCCCACTCAAGAGACGACCA 4544
Qy 1350 rIleAspGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTy 1370
Db 4545 CATCCCGGGTCCATACACAAAGAGATCCCTCGTCTCTAGTGAAGACAAAGAGACTA 4604
Qy 1370 rLeuArgArgGluAlaIleLysLeuLysArgGlnGlyIleThrProProProProProSe 1390
Db 4605 CCGGCTCGGAGGACCAAGCTCTTAAAGGGAGGAGCAAGCTCCGCCCCCAAGGCTCTC 4664
Qy 1390 rArgAspLeuThrGluAlaTyrIleThrGlnAlaLeuGlyProLeuLysLeuLysProAl 1410
Db 4665 ACGGGACCTTGACCGAGGCTTACAAAGACGAGGCTCGGGCCCCCTAAGCTGAAGCGGCG 4724
Qy 1410 rHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGlnIleProAr 1430
Db 4725 CCATGAGGGCTTGTTGGCCACGGGTGAAGAGGGGCGGCTCATCATGATGATCCCGGCG 4784
Qy 1430 gGluGluLeuArgHisIleThrProGluLeuProLeuAlaProArgProLeuLysGlnI 1450
Db 4785 CGAGGAGCTGCGGACACAGCCGAGCTGCCCTCGGCGCGCGCTCAAGAGAGGCGCTC 4844
Qy 1450 rIleThrGlnGlnIleThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerIle 1470
Db 4845 CATCAGCAGGGGACCCGCTCAAGTACGACACCGGCGGCTCCACCATGGCTCCAAAAA 4904

Qy 1470 rHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLe 1490
Db 4905 GCACGAGCTACGCTCTCTCATCGGACAGCCCGGCGGAGCTTCCACCCGTGACCCGCT 4964
Qy 1490 uAspValMetValAspAlaArgAlaLeuGluAArgAlaCysTyrGlnGluSerLeuLysSe 1510
Db 4965 GATGTGATGGCCGAGCGCCCGGACACTGAACTGCTGCTTACAGAGAGAGCTGAAGAG 5024
Qy 1510 rArgProGlyIleThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIle 1530
Db 5025 CCGGCAAGGACCGCAGACGCTCGGGGGCTCATTTGGCGGGCGCCCGGTGATTTGT 5084
Qy 1530 lProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPh 1550
Db 5085 GCCTGAGCTGGGACAGCGCGGAGAGAGCCCTTACCTATGAGAGACACCGGGACACCTT 5144
Qy 1550 rAlaGlyHisLysProArgGlySerProValThrMetArgGluProThrProArgLeuG 1570
Db 5145 TGCCGCGACCTCCACAGAGTTCCGCCGTGACCAAGCGGAGCCACCGCGGCTGCA 5204
Qy 1570 ngIuGlySerLeuSerSerSerIleValAserGlnAspArgLysLeuThrSerThrProAr 1590
Db 5205 GAGAGGACAGCTTTCCTCCAGAGGACATCCAGAGCCGAAGCTGACGTGACGCTCG 5264
Qy 1590 gGluIleAlaLysSerProHisSerThrValProGluHisIleProHisProIleSerPr 1610
Db 5265 TGAATCGCAAGTCCCGGACAGCACCGTGGCCGAGACACCAACCCCATCTTGCC 5324
Qy 1610 rTyrGlnHisLysLeuArgGlyValIleSerGlyValaAspLeuTyrArgSerHisIleProLe 1630
Db 5325 CATATGACACTGCTTCGGGGCGTGAAGTGGCGTGAACCTGTATCGACGACATCCCTC 5384
Qy 1630 uAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaIleTyrTy 1650
Db 5385 GGCCTTGACCCCACTCATACCCCGGACATCCCTGAGAGCGAGCGCTCCACTA 5444
Qy 1650 rLeuProArgHisLysAlaProAspProThrTyrProHisLysTyrProProTyrLeuI 1670
Db 5445 CCGGCCCAACCTTGCCCGGACACCGCTTACCCGACCTTACCCACCTTCACTCAT 5504
Qy 1670 rArgGlyTyrProAspThrAlaAlaLeuGluAsnArgIleIleHisAspTyrI 1690
Db 5505 CCGGCGCTACCCGACACAGCGCGCGCTGGAGAACCGGACCATCATCATCATCATCAT 5564
Qy 1690 rThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLe 1710
Db 5565 CACTTGACAGATGACACACAAACCGGCGCACCGGCATGGCCACAGAGCTGATATGCT 5624
Qy 1710 uArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgG 1730
Db 5625 GAGGGGCTCTGCCCCCGGAGATCTCTGCTGGCACTCAACTAGCGTGGGATCCCGAGG 5684
Qy 1730 rIleIleAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyTh 1750
Db 5685 CATCATGCACTGTCCCAAGTGACACCTGCTGTGTGTGCGCCCGACACAGGAC 5744
Qy 1750 rProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheserSe 1770
Db 5745 CCAAGCACCGGACATGAGACGGCTTGCTTACCTCCCAACCGGCGCCCAAGCTTCAAGCAG 5804
Qy 1770 rArgHisSerSerSerProLeuSerProGlyGlyProThrHisLysThrLysProThrTh 1790
Db 5805 CCGCACAGACGCTCCCACTTCCCAAGAGGTCCAAACACTTACAAACCAACCAAC 5864
Qy 1790 rThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluAr 1810
Db 5865 CAGCTCTCTGTCGAGAGCGGAGAGAGACGGGATGAGAGCGGAGACCGGAGATCGGAGCG 5924
Qy 1810 gGluLysSerIleLeuThrSerThrThrThrValGlnHisAlaProIleTyrArgProG 1830
Db 5925 GAAAAAGTCAATCTTCATGTCACCAAGAGGTGAGACGAGACCATCTTGAGAGACTGG 5984

QY 1830 YThrGlnInsSerSerGlySerSerGlyGlyGlyGlySerSerSerAr 1850
DB 5985 TACAGAGCAGAGCAGCGGAGAGCGGAGCGCGGGGTGGGGGAGCAGACGCCG 6044
QY 1850 GProAlaSerHisSerHisAlaHisGlnHisSerProHisSerProHisGlnAspAl 1870
DB 6045 CCCCCCTCCCACTCCCACTGCCACAGACTCGCCATCTCCCTCCGAGCCAGAGATGC 6104
QY 1870 AleuGlnGlnArgProSerValLeuHisAlaSerHisGlyMetCysGlyLeileThrAlaAva 1890
DB 6105 CCTCCAGAGAGAGACCAAGTGTCTTCAACAACAGGCAATGAAGGATATATACCGCTCT 6164
QY 1890 GlnuProSerLysProHisValLeu-----ArgSerHisSerHisSerProVa 1907
DB 6165 GAGGCCAGCAGACGCCACGGCTCTTGAGGTGGCCAGGTCCACTCCACTCTCCACCGCT 6224
QY 1907 LArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyYThrLeuAspG 1927
DB 6225 TCGCCCGGTCCCAATCCCACTGCCACCACTGCCACTGGGGGCGACCTCGATGG 6284
QY 1927 ValYThrProThrLeuMetGluProValLeuLeuProLysGlnAlaProArgValAlaAr 1947
DB 6285 GGCTACCTACCTACCTACAGAGCCCGCTTGTCTGCCAAGAGGCCCGCCGGGTGGCCG 6344
QY 1947 GProGlnuArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSe 1967
DB 6345 GCCAAGAGCGGCCCGAGCAGACCGGCCATGCTCTCCGCCAAGCCCGCCGCGCTC 6404
QY 1967 GlnYLeuGlnuProAlaSerSerProSerLysGlySerGluProArgProLeuValProPr 1987
DB 6405 CGGGGTGAGGCCCGCTCTCCCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6464
QY 1987 CValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisAla 2007
DB 6465 TGTCTCTGGCCACGCGCACCATGCGCCGACCCCTGGAGAAACCTCGACCTCACCACGC 6524
QY 2007 AserProAspProProAlaProProAlaSerAlaSerAspProHisArgGlnuLysThrG 2027
DB 6525 CAGCCCGGAGCCCGCGGCGCACCTGGCTCGGCTCGGAGCCGCGGAGAAAGACTCA 6584
QY 2027 nSerLysProPheSerIleGlnGlnuLeuGlnuLeuArgSerLeuGlyYThrHisGlySerSe 2047
DB 6585 AAGTAAACCTTTTCATCCAGAACTGGAACTCGGCTTCTGGGTTACACAGCGCAGCAG 6644
QY 2047 TYrSerProGlnuGlyValGluProValSerProValSerSerProSerLeuThrHisAs 2067
DB 6645 CTACAGCCCCGAGAGGGGTGAGCCCGCTCAGCCCTGTGAGCTCACCAGCTACACCA 6704
QY 2067 PLYSGLYLeuProLysHisLeuGlnuGlnuLeuAspLysSerHisLeuGlnuGlyGlnuLeuAr 2087
DB 6705 CAAGGGGCTCCCAAGCAGCTGGAAGAGCTCGCAAGAACCACTGAGGGGAGGAGCTGCG 6764
QY 2087 GProLysGlnProGlyProValLysLeuGlyGlyValAlaHisLeuProHisLeuAr 2107
DB 6765 GCCCAGAGCAGCGCCCGGTGAGGTCGGGGAGGCCGCCACCTCCCACTCCGTCGG 6824
QY 2107 GProLeuProGlnuSerGlnProSerSerSerProLeuLeuGlnuThrAlaProGlyValLys 2127
DB 6825 GCCGCTGCTGAGAGCCAGCCCTCTGTCCAGCCCGCTGTCCAGACCGCCCGAGGGGTCAA 6884
QY 2127 sGlnHisGlnArgValValThrLeuAlaGlnHisIleSerGlnuValIleThrGlnAspTY 2147
DB 6885 AGGTACACGAGCGGTGTGTCACCTGGCCAGCAGCATGAGGTATCATCACAGAGACTA 6944
QY 2147 rThrArgHisHisProGlnGlnuLeuSerAlaProLeuProAlaProLeuTYrSerPhePr 2167
DB 6945 CACCCGGCAGCAGCAGCAGCAGCTCAGCCAGCCCGCCCGCCCTTACTCTCTCC 7004
QY 2167 oGlnYAlaSerCysProValLeuAspLeuArgArgProProSerAspLysLeuTYrLeuProPr 2187
DB 7005 TGGGGCCAGCTGCCCGTCTGAGACTTCGGCGGCCCAACCAAGTACCTTACTCCGCGC 7064
QY 2187 oProAspHisGlnAlaProAlaArgGlySerProHisSerGlnuGlyLysArgSerPr 2207

DB 7065 CCGGAGCATGATGCCCCGCGCGCTCCCGCCCAAGCGGAGGGGCAAGAGTCTCC 7124
QY 2207 oGlnuProAsnLysThrSerValLeuGlyGlyGlyLysAspGlyIleGluProValSerPr 2227
DB 7125 AGAGCCAAACAGACGCTGGCTTGGGTGTGTGAGAGCGGATATGAACTGTGTCTCCC 7184
QY 2227 oProGlnuGlyMetThrGluProGlyHisSerArgSerAlaValTYrProLeuLeuTYrAr 2247
DB 7185 ACCGAGGGCATGACGAGAGCGGAGCTCCCGGAGTGTGTATCCGCTGTCTATCCG 7244
QY 2247 GAspGlyGlnuGlnuThrGluProSerArgMetGlySerLysSerProGlyYAsnThrSerG 2267
DB 7245 GGATGGGAGAACAGAGAGGCCAGAGATGGGCTTCGAAGTCTCCAGGCAGAACAGCACA 7304
QY 2267 nProProAlaPhePheSerLysLeuThrGlySerAsnSerAlaMetValLysSerLysLys 2287
DB 7305 GCCGCCAGCTTCTTTCAGCAAGCTACCGAGAGCACTCCGCAATGTCTCAAGTCCAGAA 7364
QY 2287 sGlnuGlnuLeuAsnLysLysLeuAsnThrHisAlaAsnArgAsnGluProGluTYrAsnIleSe 2307
DB 7365 GCAAGAGATCAACAGAAAGCTGAACACCCCAACCGGAATGAGCTGAAATACAAATATCAG 7424
QY 2307 GlnProGlyYThrGlnuIlePheAsnMetProAlaIleThrGlyYThrGlyLeuMetThrTY 2327
DB 7425 CCAGCTTGAGACGGAGATCTTCAATATGCGCCATCAACGGAAACAGGCTTATGACCTA 7484
QY 2327 rArgSerGlnAlaValGlnGlnuHisAlaSerThrAsnMetGlyLeuGlnuAlaIleAr 2347
DB 7485 TGAAGCCAGCGGCTGCAAGAACATGCCAGCAACAACTGGGGCTGAGAGCCATATATG 7544
QY 2347 GlnYAlaLeuMetGlyLysTYrAspGlnTYrGlnuGlnuSerProProLeuSerAlaAsnAl 2367
DB 7545 AAAGGCACTCATGGGTAAATATGACCAATGGGAAAGTCCCGCGCTCAGCGCCAATGC 7604
QY 2367 aPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlAs 2387
DB 7605 TTTTAACTCTGAAATGCGACAGCGCAGCTGCGCTGTATGCCATPAACGCTGTGA 7664
QY 2387 PGLYArgSerAspHisSerLeuThrSerProGlyGlyGlyLysAlaLysValSerG 2407
DB 7665 CGGAGGAGTACCAACACTACTCCGCAAGTGGCGGGGAAAGCCAAAGTCTCTGG 7724
QY 2407 YArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgPr 2427
DB 7725 CAGACCCAGACAGCGAAAGCCAAAGTCCCGGCGCTCGGATCTGGGAGACCGGCC 7784
QY 2427 oProSerValSerSerValHisSerGlnuGlyAspCysAsnArgArgThrProLeuThrAs 2447
DB 7785 ACCCTCTGTCTCTCAGTGACCTCGAGGAGACTGCAACCGCGAGACCGCTCACTCA 7844
QY 2447 nArgValTYrGlnuAspArgProSerSerAlaGlySerThrProPheProTYrAsnProLe 2467
DB 7845 CCGCTGTGTGGAGAGCAGGCGCTCGTCCGAGGTTCCAGCCATTCCTTACACCCCT 7904
QY 2467 uIleMetArgLeuGlnuAlaGlyValMetAlaSerProProProProGlyLysProAlaG 2487
DB 7905 GATCATGGCGGTGCAAGCGGGGTGTATGCTTCCCAACCCCAACCGGGGCTCCCGCGGG 7964
QY 2487 YSerGlyProLeuAlaGlyProHisHisAlaThrAspArgGlnuProLysProLeuLeuY 2507
DB 7965 CAGCGGGCGCTCGCTGGCGCCCAACGCTGGGAGCAGAGGCCCAAGCCACTGCTGTG 8024
QY 2507 sSerGlnuTYrGlnuThrLeuSerAspSerGlu 2517
DB 8025 CTTCGAGTACGAGACACTCCGACAGCGAG 8055

RESULT 15
ADQ23294/c
ID ADQ23294 standard; DNA; 9079 BP.
XX
XX
AC
XX ADQ23294;

DT 26-AUG-2004 (first entry)
XX Human soft tissue sarcoma-upregulated DNA - SEQ ID 6114.
XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KM db.
XX Homo sapiens.
XX MO2004048938-A2.
XX
XX 10-JUN-2004.
XX
XX 26-NOV-2003; 2003WO-US038193.
XX PF
XX 26-NOV-2002; 2002US-0429739P.
XX
XX (PROT-) PROTEIN DESIGN LABS INC.
XX
XX Aziz N, Ginsburg WM, Zlotnik A;
XX WPI; 2004-441208/41.
XX
XX
XX
XX
XX Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
XX
XX Example 2; SEQ ID NO 6114; 210pp; English.
XX
XX The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
XX
SQ Sequence 9079 BP; 1268 A; 2703 C; 3184 G; 1924 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 0 Length: 9079
Score: 12845.50 Matches: 2473
Percent Similarity: 98.10% Conservative: 6
Best Local Similarity: 97.86% Mismatches: 20
Query Match: 97.20% Indels: 31
DB: 12 Gaps: 3
US-09-522-753-5 (1-2517) x ADQ23294 (1-9079)
QY 1 MetSerG1SerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro 20
Db 8552 ATGTCCGGGCTCCACACAGCTGTGGACACAGCGTGGAGGCGACTGAGCCCCGCTACCCG 8493
QY 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
Db 8492 CCCACAGGCTTTCCTAACCAAGTGCAGATGCGCCGACGACACGACGTCGGGCTCTCG 8433
QY 41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
Db 8432 GAGTACACAGACACACTCCGCGACTATGCTCCACACTGTGCGCCGCTCCATCATCCAG 8373
QY 61 ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyIleGlnIleArgSerGln 80
Db 8372 CCCACAGGCGGAGGCGCTCTCCCTGCTGTCTGAGTTCCAGCCCGGAAATGACGCTCCAG 8313
QY 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyIleSerGluMet 100

Db 8312 GAGCTCACCTGCGGCGCAGAGTCCCACTCATACCTGCCGAGCTGGGAACTCAGAGATG 8253
QY 101 GluPheIleGluSerTyrArgProArgLeuGluLeuProArgProLeuLeuArgPro 120
Db 8252 GAGTTCAATTGAAACCAAGCCGCTCGCTTAAGCTGCTGCTGACCCCTCTCTGGACCG 8193
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrThyAspArgSer 140
Db 8192 TCACCCCTGCTGGCCAGCGGCGACCTGCGGGACTGTGAAGACTTCACCAAGACCGGCC- 8134
QY 141 LeuThrGlyLeuLeuGluProValSer-ProProSerProProHisThrAspProGluLe 160
Db 8133 CTGACGGGCAAGCTGGAACCGGTGTCTCCACCCGACCCCGCACACTGACCTCTGAGCT 8074
QY 160 UGULeuValProProArgLeuSerTyrGluLeuIleGlnHisMetAspArgValAs 180
Db 8073 GGAAGCTGTGCGCGCAGCGCTGTCCAAAGGAGGAACTATCCAAACATGACCCGCTGGA 8014
QY 180 PArgGluIleThrMetValGluGlnGlnIleSerTyrLeuValTyrValGlnGlnIle 200
Db 8013 CCGAGAGATCACCATGTGTAGACGACGATCTTAAGCTGAAGAAAGACGACACAGCT 7954
QY 200 UGUUGUUGUAlaAlaLeuSerProGluProGluIleValSerProProIleG 220
Db 7953 GAGAGAGAGAGCTGCGCAAGCGCGCGAGCTGAGAAACCGGTGTACCGCGCCCATCGA 7894
QY 220 USerTyrHisArgSerLeuValGlnIleIleTyrArgGluAsnArgTyrValAlaGluAl 240
Db 7893 GTGAAACACCGCAGCTGTGTGAGATCATTCAGACGAGAAACCGAAGAGCTTGAGC 7834
QY 240 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPr 260
Db 7833 TGCACATCGAATTTGGAAGGCTTGAGGCCCCCAGGTGAGTGGCTGCTGTACACAGGCC 7774
QY 260 OSerAspThrArgGlnTyrHisGluAsnIleTyrIleAsnGlnAlaMetArgTyrValLe 280
Db 7773 CTCGACACCCGCGAGATCATGAAACATCAAAATTAACACAGCGATGCGAAGAAAGCT 7714
QY 280 UileLeuTyrPheLeuArgArgAsnHisAlaArgTyrGlnTyrValGlnTyrPheCysG 300
Db 7713 AATCTGTACTTCAAGAGGAGGAATACGCTCGGAAACAATGGAGCAGAAATTTCTCCA 7654
QY 300 nArgTyrAspGlnLeuMetGluAlaLeuGluIleValGlyValGluArgIleGluAsnAsnPr 320
Db 7653 GCGCTATGACCACTCATGAGGCTCTGGGAAAGAGTGAAGCGATTCAGAAACAAACC 7594
QY 320 oArgArgArgAlaLeuGluSerTyrValArgGluTyrTyrGluValGluAsnProGluI 340
Db 7593 CCGGCGCGCGGCGCAAGAGAGCAAGGTTCGAGTACTACAGAAAGCAGATTCCTGAGAT 7534
QY 340 eArgTyrGlnArgGluLeuGlnGluIleValGlyMetGlnSerArgValGlyGlnArgGlySerG 360
Db 7533 CCGCAAGACGCGGAGCTGTGACGAGCGCATGCGAG--AGGCTGTGCGCAGCGGCGCACTGG 7477
QY 360 yLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSe 380
Db 7476 GCTGTCTCATGTCCCGCGCGCGCGACGACGACGAGGTCTCAGAAATCATCATGAGCTCTC 7417
QY 380 rGluGlnGluAsnLeuGluValGluGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400
Db 7416 AGAGCAGAGAACTGTGAAGAAAGATGCGCGCAGCTGCGCGTATCCCGCCCATGCTGT 7357
QY 400 rAspAlaAspGlnGlnArgIleLeuValPheIleAsnMetAsnGlyLeuMetAlaAspProMe 420
Db 7356 CGACGCTGACGACGACGACATCACTTCATCAACATGAAGAGGCTTATAGCCACCCCAT 7297
QY 420 tLysValTyrTyrAspArgGlnValIleMetAsnMetTPSerGluGlnGluValGluThrPh 440
Db 7296 GAAAGGTGTAACAAGACCGCAGGTCTATGAACATGTGAAGTGAAGCAGAGAAAGAGACCTT 7237
QY 440 eArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleIleSerPheLeuGluLar 460

Db 7236 CCGGAGAAAGTTTCATGCAGCATCCCAAGAACTTTGGCTGATGCATCTTCCTGGAGAG 7177
QY 460 GlycerThrValaGluCysValLeuTyrTyrTyrLeuThrIleValAsnGluAsnTyrIle 480
Db 7176 GAAGCAGGCGCTGAGTGGCTCTCTATTATCACTGACGAAAGAAAGAAACATATTA 7117
QY 480 sSerLeuValaGAGAsSerTyrArgArgArgIlyIySerGlnGlnGlnGlnGlnGlnGln 500
Db 7116 GAGCTGGGTGAGACGGAGCTATCGGCGCGCGCAAGAAC--CAGCACAAACAAAG 7059
QY 500 ngInGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
Db 7058 CCGGAGGAGCAGCAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGAA 6999
QY 520 sAspGluValSerGluValSerGluValaGluValSerGluValaGluValaGluVala 540
Db 6998 AATATAGAAAGAGAAAGAAAGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 6939
QY 540 nAspIySerGluAspLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 560
Db 6938 CGACAAAGAAAGACTCTCTCAAGAGAAAGACAGACACTCAAGGAGAGACAGACAGCA 6879
QY 560 uIySerGluValaValaIaSerIySerIyArgIyThrIaAsnSerGlnIyArgArgIySer 580
Db 6878 GAAGGAGGCTGTGGCTCTCAAGAGAGCGCAAAACTGCAACAGCCAGAGAGAGAGAGAG 6819
QY 580 yArgIleThrArgSerMetIaAsnGluIaAsnSerGluIaIleThrProGlnGln 600
Db 6818 CCGCATCACCCGCTCATATGATGAGGCACACAGAGAGAGGCAATCAACCCCAAGCA 6759
QY 600 nSerIaGluLeuIaAsnMetGluLeuAsnGlnIySerArgTyrThrGluGluGluGlu 620
Db 6758 GAGCGCGCAGACTGGCTCTCATGAGAGCTGATGAGATTCCTCGCTGACAGAAAGAAAT 6699
QY 620 rGluThrIaIyIyIySerGluLeuLeuGluIyIyIyIyArgAsnTyrPheAsnTyrIyIyIy 640
Db 6698 GGAACAGGCAAGAAAGGTCTCTCTGAGAACAGCGCCCAACTGGTCCGCCATCCCGGAT 6639
QY 640 tValGlySerIyThrValSerGlnCysIyAsnPhetIyPheAsnTyrIyIyIyIyIyIy 660
Db 6638 GGTGGGCTCCAGAGCTGTGTGCGAGTGTAAAGACTTACTTCACTCAAGAGAGAGCA 6579
QY 660 nAsnLeuAspGluIleLeuGlnGlnIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 680
Db 6578 GAACCTCGATGAGATTTGAGAGACAGCAACAGCTGAAGATGAGAGAGAGAGAGAGAG 6519
QY 680 gArgIy 700
Db 6518 GAGGAAG 6459
QY 700 lGluAspGluGluMetGluIaSerGluIyIaSerGluIyAsnGluGluMetValGluGlu 720
Db 6458 GAGAGATGAGAGATGAG 6399
QY 720 uIaGluIaIleuHiaIaSerGluIyAsnGluValProArgIyGluCysSerIyProAl 740
Db 6398 GCGTGA-----GC 6390
QY 740 aThrValaAsnAsnSerSerAspThrGluSerIleProSerProHiaIThrGluIaIyIy 760
Db 6389 CACTGTCAACAAACAGCTCAGACACCAAGAGCATCCCTCTCTCACTCACTGAGGCCCA 6330
QY 760 sAspThrGluIyAsnGluIyProIyAspProAlaThrLeuGluValaAspGluProProP 780
Db 6329 GAGACAGGCGCAAAATGGGCCCAAGCCCAAGCCCACTGGGGCGCGAGCGGAGCAACCCC 6270
QY 780 ogIyProProThrProPro-ArgArgThrSerArgIaProIleGluProThrProAlaAs 800
Db 6269 AGGGCCACCAACCCCAACACCGAGAGACATCCGGCC-CCCACTGAGTCAACCCCGGCT 6211
QY 800 eArgIuIaIy 820
Db 6210 CTGAAGCCACCTTAGCCCTTACGCCCCCAAGACACCCCAATTCCTCTTCACTCTCCTC 6151

QY 820 roValIaIyProIySerGluGluIySerGluGluGluThrIaIaIaIaProProValIyGluGlu 840
Db 6150 CTGTGTCCTCCCAAG 6091
QY 840 lYglGluGluIyIyProProAlaIaGluGluLeuValaIaIyIyIyIyIyIyIyIyIyIy 860
Db 6090 GGGAG 6031
QY 860 lProValIySerSerGluCysThrGluGluIaIyGluGluIyProAlaIyIyIyIyIyIy 880
Db 6030 AGCCGCTCAAG 5971
QY 880 lAGluIaIaIaGluIaIy 900
Db 5970 CCGAGGCGCTAGGCGCAAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5911
QY 900 eArgIyArgIaIy 920
Db 5910 CCGGAGAGGCGCAACAGCAG 5851
QY 920 hrcYSerIaIaAspGluValaAspGluIaGluGluIyIyIyIyIyIyIyIyIyIyIyIy 940
Db 5850 CCGAGAGTGCAG 5791
QY 940 rArgProSerLeuLeuThrProThrGluIyAspProArgIaIaAsnIaSerProGlnIyAsp 960
Db 5790 CAAGGCCAGCTCTCTCAACCCGAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5731
QY 960 rLeuAspLeuIySerGlnLeuIy 980
Db 5730 CACTGAGCTTGAAG 5671
QY 980 yValIy 1000
Db 5670 AAGTTCATAG 5611
QY 1000 rProProGlnIyAsnLeuGlnProGluSerAspIaIyIyIyIyIyIyIyIyIyIyIyIy 1020
Db 5610 CCGACCGCAAAACCTGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5551
QY 1020 rGluIyIySerIySerProAlaProProAlaIaAspIyIyIyIyIyIyIyIyIyIyIy 1033
Db 5550 GGGGCAAG 5491
QY 1034 -----AlaPheIaIaIaGluIaIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 1052
Db 5490 TCCAGCTTCGAGCGAG 5431
QY 1052 lYleuProPheProValProProArgIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 1072
Db 5430 GCTGTCCTTCCCGAG 5371
QY 1072 rSerIaIaPheSerIyIaIaProProGluIyIaIyIyIyIyIyIyIyIyIyIyIyIyIy 1092
Db 5370 CCTCAGCTTCTCTCTAGCTTCACTGCTCAGCCTGCTCAGCCTGCTCAGCCTGCTCAG 5311
QY 1092 lAspProValIy 1112
Db 5310 CCGGCGCGCTTCCGCGAG 5251
QY 1112 lAlaIy 1132
Db 5250 CCAAGCAGCCAG 5191
QY 1132 lIy 1152
Db 5190 AGCTTCAGCTTCCGCTTCAAG 5131
QY 1152 eAspProIy 1172
Db 5130 TGCCTCTGCGCATGAG 5071

QY 1172 euserProatrglyglnalaglyProProgiuserleuglyvalProthralaglnlva 1192
DB 5070 TGTCCTCCAGCGGCGCCAGGCTGGGCCACCGGAGAGCCTGGGGGTCCCAACGCCAGGAGG 5011
QY 1192 laSerValleuarglylThralaleuglySerValProgiyglySerlIethrlysglyt 1212
DB 5010 CGTCGGCTGAGAGGAGCAGCTCTGAGCTCAgTTCGGGGGAGAGCATCCAAAGCA 4951
QY 1212 leProSerThArSValProSerApsSerAlaIethrThYArgGlySerlIethrhias 1232
DB 4950 TTCCACACACAGCGGTCCCTCCGACAGCCCATCACTACCGGGCTCCATCCACAG 4891
QY 1232 lYthProalAapValleuThYrlysglylThrlIethrargylelsglyglnuapserp 1252
DB 4890 GCACGCGAGTGAAGCTCTGTACAAAGGACCATCACAGGATCATCGGCGAGGAGACC 4811
QY 1252 roSerArGleuAapArGlyArGlyuapSerleuProlyglnvalIlelyrGluG 1272
DB 4830 CGAGTCGCTTGACCGCGCGGAGGAGCAGCTGCCCAAGGGCCAGCTCATACGAG 4771
QY 1272 lYlYalYsglyHlSvalleuSerThYrGlyGlyMetSerValThrGlnCySerlysg 1292
DB 4770 GCMAAGAGGCGCAGCTTGTCTCTATAGAGGTGACATGTCTGACCCAGTCTCAAG 4711
QY 1292 luAapGlyArGSerSerSerglyProProhIsgIuthralaIaProlysaThrTYrA 1312
DB 4710 AGGACGCGAGAGAGCTCAGAGCCCCCATGAGAGCGCGGCCCAAGGACCTATG 4651
QY 1312 sPmetMetGlyGlyArGValGlyArGAlaIleSerSerAlaSerlIleGlyGlyleuMetG 1332
DB 4650 ACATGAGAGGAGCGCGGTGGGAGAGCATCTCTAGCCAGCATGAAGGTCTCATGG 4591
QY 1332 lYArGAlaIleProProgiuArGHisSerProhIshIleuLyGlnGlnHlshIleA 1352
DB 4590 GCGGTGCTCATCCGCGGAGGAGCAGACCCCACTCTTAAGAGCAGACCATCACTCC 4531
QY 1352 rglySerlIethrGlnGlylIeProArGSerThYrValGlnlAglnlUaSPlyrleuA 1372
DB 4530 GCGGCTCATCACACAAGGATCCCTCGTCTATCGTGAAGGACAGAGAGACTACTGC 4471
QY 1372 rGArGlyuAlaYleuLeuLyArGlyGlnGlylThrProProProProProSerArG 1392
DB 4470 GTGCGGAGGCGCAAGCTCTTAAGGAGGAGGAGCGCTCCGCCCAACCCCTCACGGG 4411
QY 1392 sPleuThrGlnAlaYrlyrlyeThrGlnAlaLeuGlyProleuLySleuLySProAlaHiasG 1412
DB 4410 ACCTGACCGAGGCTACACAGCAGGCGCTGGGCCCCCTGAAGCTAAAGCCGCGCATG 4351
QY 1412 lUglyleuValaIaThrVallysglyuAlaGlyArGSerlIehIsgIuIleProArGlyug 1432
DB 4350 AGGCGCTGGTGGCCAAGTGAAGAGGCGGCGCTCATCATGAGATCCCGCGGAGG 4291
QY 1432 luleuArGHisThrProgiuLeuProleuAlaProArGProleuLyGlyGlySerlIeT 1452
DB 4290 AGCTGGGCGCAGCGCGGAGCTGCCCTGGCCCCCGCGCTCAAGAGAGGCTCCATCA 4231
QY 1452 hrcGlnGlyThrProleuLyArSPTThrGlyAlaSerThrThrGlySerlySleuLyshIa 1472
DB 4230 CGAGAGGACCGCGCTCACTAGTACGACCGCGGCTCCACACTGGTCCAAAGAGCAG 4171
QY 1472 sPValArGSerleuIleGlySerProGlyArGThrPheProProValHlaSProleuApy 1492
DB 4170 ACCTAGCTCTCATCTGCGAGCCCGGCGGAGCTTCCACCGGTCACCGCTGGATG 4111
QY 1492 alMetAlaAapAlaAaGAlaLeuGlnArGAlaCybTyrglyGlnuSerleuLySerArGp 1512
DB 4110 TGAATGCGCAGCGCGGAGCTGGAACGTGCTGTACGAGAGAGCTTAAAGCGCGC 4051
QY 1512 roGlylThralaSerSerSerglyGlySerlIeAlaArGlyAlaProValIleValProG 1532
DB 4050 CAGGAGCGGCGCAGAGTGGGGGCTCCATTGCGCGGCGCGCGCTCATTTGTGCTG 3991
QY 1532 luLeuGlyLySProArGlnSerProleuThrTYrGlnuAapHlsglyAlaProPheAlaG 1552

DB 3990 AGCTGGGTAGCCGCGGAGAGCCCTGACCTATGAGAGCACAGGGGACCCCTTTCGG 3911
QY 1552 lYHlSleuProArGlySerProValThrMetArGlyuPProThProArGlyGlnlG 1572
DB 3930 GCACCTTCCACAGAGTTCGCCGTGACATGCGGAGGCCACCGCGCTGAGAGG 3871
QY 1572 lYserleuSerSerSerglyAlaSerGlnAapArGlySleuThSerThProArGlyI 1592
DB 3870 GCAGCTTGTGTCCAGAGAGCATCCAGAGACGGAAGCTGACTGACAGCTGTGAGA 3811
QY 1592 leAlaLySerProhIshSerThrValProgiuHlshIshProhIshProIleSerProTYrG 1612
DB 3810 TCCCAAGTCCCGCAGAGCAGCAGCTGCCAGAGACCAACACCCCATCTCGCTATG 3751
QY 1612 luHlSleuAarglyValSerGlyValaApleuThYrArGSerHlIleProleuAlaP 1632
DB 3750 AGCAGCTGTCCGGGGGTAGTGGCTGGACCTGTATCGACGCACATCCCTGGCT 3691
QY 1632 heAapProThSerlIeProArGlylIeProleuAapAlaAlaIaTYrTYrleuP 1652
DB 3690 TCGACCCCATCTCATACCCCGGCGATCCCTTGAGCGAGCGCGCTTACTACTGC 3631
QY 1652 roArGHisleuAlaProAapProThrTYrProhIshleuTYrProProTYrleuIleArG 1672
DB 3630 CCCGACACTGGCGCCCAACCCCACTACCGGACCTGTACCACTCATCCGCGG 3571
QY 1672 lYTYrProAapThralaAlaLeuGlnuAapArGlnThrlleIleAapArTYrIleThrs 1692
DB 3570 GCTAACCCGACAGCGCGCGCTGAGAACGAGACATCATCAATCACTACTCACT 3511
QY 1692 eGlnGlnMetHlshIshAenThralaThralaMetAlaGlnAlaAapMetleuArG 1712
DB 3510 CGAGACATGACACACACACAGCCGACCTGGCCAGAGCTGATGTGAGG 3451
QY 1712 lYleuSerProArGlyuSerSerleuAlaLeuAnTYrAlaAlaGlyProArGlylIeI 1732
DB 3450 GCTCTCGCCCGCGAGTCTCGCTGCACTCAACTACGCTGGGGTCCCGAGGATCA 3391
QY 1732 leApleuSerGlnValProhIshleuProValleuValProProThProGlylThrProA 1752
DB 3390 TCGACCTGTCCAAAGTGCACACCTGCTGTGTCGCCCCGAGACACAGGACCCAG 3331
QY 1752 laThralaMetAapArGlyleuAlaTYrleuProThralaProgiuProPheSerArG 1772
DB 3330 CACCGCGCATGACCGCTTGTCTACTCCCAACCGCGCCGCTTACAGAGCGGCC 3271
QY 1772 isSerSerSerProleuSerProGlylYProThrhIshleuThlySProThThThrs 1792
DB 3270 ACAGAGCTTCCCACTCTCCCGAGAGGTCAACACACTGACAAACCAACCAACGCT 3211
QY 1792 eTserSerGlyuArGlyuArGArGArGlyuArGArGArGArGArGArGlyuArGlyuL 1812
DB 3210 CTTGTCGAGCGGAGGAGCAGACCGGATGCAAGGAGGAGCGGATCGGGAACGGGAAA 3151
QY 1812 ySserlIleuThrSerThrThThrThValGlnHlshIaProlIeTPArGProGlyThrG 1832
DB 3150 AGTCATCTCTCACTCAACACGACGAGTGAAGACCAACCATCTGAGACTGTGACG 3091
QY 1832 lUglnSerSerglySerSerglySerSerglyGlyGlyGlyGlyGlySerSerArGProA 1852
DB 3090 AGCAGAGCAGCGGACAGAGGAGCAGCAGCGCGGGGTGGGAGCAGAGCGCGCCG 3031
QY 1852 laSerHlshSerHlshAlaHlshGlnHlshSerProIleSerProArGThrGlnuAapleuG 1872
DB 3030 CTTCCCATCTCCATGCGCACAGACTCGCCATCTCTCCCTGGAGCCAGAGTCCCTCC 2971
QY 1872 lnglnArGProSerValleuHlshAenThrGlyMetlysglylIeIleThralaValGlnP 1892
DB 2970 AGCAGAGACCGAGTGTGCTTCAACACAGGATGAAGGTATCATACCGCTGTGAGC 2911
QY 1892 roSerlySProThrValleuArGSerThSerThSerSerProValaProAlaAlaT 1912

Db 2910 CCAGAAAGCCAGGCTCTGAGGTCCACTCCACTCTGACCCGTTGCGCAGCTGCCA 2851
Qy 1912 hPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValIYrProThrL 1932
Db 2850 CATTCCCACTGCTCCACCACTGCTCCACTGCGCGGCACTCTGATGGGGCTCTACCC 2791
Qy 1932 eumEGluProValLeuLeuProLYGluAlaProArgValAlaArgProGluArgProA 1952
Db 2790 TCATGGAGCCCGCTTGTGCTGCCAAGAGAGGCCCGCGGGTCCGCCGAGAGCGGCC 2731
Qy 1952 rGAlAspThrGlyHisAlaPheLeuAlaLYsProProAlaArgSerGlyLeuGluProA 1972
Db 2730 GAGCAGACACCGGCGCACTGCTCTGCAAGCGCCGACCGCTCCGGAGTGGAGCCCG 2671
Qy 1972 lAsSerSerProSerLYsGlySerGluProArgProLeuValProProValSerGlyHisA 1992
Db 2670 CCTCTCTCCCGCAAGAGGCTCGAGGCCCGGCGCTTATGCTCTGTCTGCGCAAG 2611
Qy 1992 lAThrTlEAlaArgThrProAlaLYsAsnLeuAlaProHisHisAlaSerProAspProP 2012
Db 2610 CCACCAATCCCGCGCACTCTGCGAAGAACCTCGCACTCACACGCGCAGCCGAGCCG 2551
Qy 2012 rGAlAProProAlaSerAlaSerAspProHisAArgGlyLeuThrGlnSerLYsProPheS 2032
Db 2550 CGGCGCACCTGCTGCGCTCGAGCCGACCGGAGAAAGACTCAAGATAACCTTTT 2491
Qy 2032 eTlEgInGluLeuGluLeuArgSerLeuGlyYrHisGlySerSerTYrSerProGluG 2052
Db 2490 CCATCCAGGAACTCGAATCTCGTTCTCTGGGTTTACACGGCAGCACTACGCGCGAAG 2431
Qy 2052 lYValGluProValSerProValSerSerProSerLeuThrHisAAspLYsGlyLeuProL 2072
Db 2430 GGGTGGAGCCCGTCAGACCTCTGAGCTCACCCAGTGTGACCCAGCAAGAGGGCTCCCA 2371
Qy 2072 YHisLeuGluGluLeuAspLYsSerHisLeuGluGlyGluLeuArgProLYsGlnProG 2092
Db 2370 AGCACTGGAAAGCTCGACAGAGCCACTGGAGGGAGCTGCGGCCCAAGACCCAG 2311
Qy 2092 lYProValLYsLeuGlyGlyGluAlaAlaHisLeuProHisLeuAspProLeuProGluS 2112
Db 2310 GCCCGGTGAAGCTTGGCGGGAGGCCGCCACTCCACACCTGCGCGCTCCCTGAGA 2251
Qy 2112 eTGlNProSerSerSerProLeuLeuGlnThrAlaProGlyValLYsGlyHisGlnArgV 2132
Db 2250 GCGAGCCCTGCTCGACACCGCTGCTCCACACCGCCAGGGGTCAAAAGTTCACACGGG 2191
Qy 2132 AlValThrLeuAlaGlnHisLeuSerGlyValLlEthrGlnAspTYrThrArgHisHisP 2152
Db 2190 TGGTCACTCTGGCCGACATCAGTAGGTCAACAAGACTACACCCGCGACACCC 2131
Qy 2152 rGInGlnLeuSerAlaProLeuProAlaProLeuTYrSerPheProGlyAlaSerCysP 2172
Db 2130 CACAGAGAGTCAAGCGAACCTGCGCGCCCTTACTCTTCCCTGGGGGCGAGCTGGC 2071
Qy 2172 iValLeuAspLeuArgArgProProSerAspLeuTYrLeuProProProAspHisGlyA 2192
Db 2070 CCGTCTCGAAGCTCCCGCCGCCACCAAGTGAAGCTTACCTCCGCGCCCGGACATAGT 2011
Qy 2192 lAProAlaArgLYsSerProHisSerGlyGlyLYsArgSerProGluProAsnLYsT 2212
Db 2010 CCGCGCGCGGTGGCTCCCCACAGGAGAGGGGCAAGAGTCTCCAGAGCCAAACAAG 1951
Qy 2212 hSerValLeuGlyGlyGlyGluAspGlyLYsGluProValSerProProGluGlyMet 2232
Db 1950 CGTGGTCTTGGTGGTGGAGAGCGATTTGAACCTGTCTCCACCGGAGGGCATGA 1891
Qy 2232 hTGlUProGlyHisSerArgSerAlaValTYrProLeuLeuTYrArgAspGlyGluGlnT 2252
Db 1890 CGAGAGCCAGGGGCACTCCCGAGGTGCTGTACCCGCTGTACCGGGATGGGGAAACA 1831
Qy 2252 hTGlUProSerArgMetGlySerLYsSerProGlyAsnThrSerGlnProProAlaPheP 2272
Db 1830 CGAGAGCCAGAGAGATGGCTCCAAAGTCTCCAGGCAACCAAGCCGCGCAAGCTTCT 1771

Qy 2272 hSerLYsLeuThrGluSerAsnSerAlaMetValLYsSerLYsGluGlnLYsEAlE 2292
Db 1770 TCAGCAAGCTGACCGAGAGCAATCCGCCATGTCAAGTCCAAAGAGATCAACA 1711
Qy 2292 YsLYsLeuAsnThrHisAsnAspGluGluProGluTYrAsnLYsSerGlnProGlyThG 2312
Db 1710 AGAAGCTGAACAACCAACAACGGAAATGAGCTGAAATACATATCAAGCCAGCTGGAGCG 1651
Qy 2312 lUlEPheAsnMetProAlaLlEthrGlyThGlyLeuMetThrTYrArgSerGlnAlaV 2332
Db 1650 AGATCTTCAATATGCTCCGCCCATCACCGGAACAGGCTTATGACTATGAACCGAGCGG 1591
Qy 2332 AlGInGluHisAlaSerThrAsnMetGlyLeuGluAlaLlEalArgLYsAlaLeuMetG 2352
Db 1590 TCGAGGAACATGCGACACCAATGAGGCTGAGGCGCATATTTGAAGGCACTCATG 1531
Qy 2352 lYLYsTYrAspGlnTrpGluGluSerProProLeuSerLYsAsnAlaPheAsnProLeu 2372
Db 1530 GTAAATATGACCAAGTGGAGAGTCCCGCGCTCGATGCGCCCAATGCTTTTAACTCTGA 1471
Qy 2372 snAlAserAlaSerLeuProAlaAlaMetProLYsEThrAlaAlaAspGlyArgSerAspH 2392
Db 1470 ATGCCAGTCCAGCTGCGCGCTGCTATGCTCATTAACCGCTGACGAGCGAGTGAAC 1411
Qy 2392 lAThrLeuThrSerProGlyGlyGlyGlyLYsAlaLYsValSerGlyArgProSerSer 2412
Db 1410 ACACACTCACTCGCGAGGTGCGGGGAGGCCAAGGCTCTGCGCAGACCCAGACGCC 1351
Qy 2412 rGlyAlaLYsSerProAlaProGlyLeuAlaSerGlyYAspArgProProSerValSerS 2432
Db 1350 GAAAGCCAAAGTCCCGGCGCGGCTGAGCATCTGGAGACGGGCCACCTCTGTCTCT 1291
Qy 2432 eTValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTTPGluA 2452
Db 1290 CAGTGCATCGGAGGAGAGCTGCMAACCGCGGACCGCGCTCAACCGCGTGGAGAG 1231
Qy 2452 sPArgProSerSerAlaGlySerThrProPheProTYrAsnProLeuLlEMetArgLYeUG 2472
Db 1230 ACAGGCCCTCGTCCGACGTTCCAGCGCCATTCCTTACAAACCCCTGATCATGGGCTGC 1171
Qy 2472 lAlAGLYsAlMetLYsSerProProProProGlyLYsLeuProAlaGlySerGlyProLeu 2492
Db 1170 AGCGGGGTCTATGGCTTCCACACCCCGAGGCTCCCGGGGACGGGCGGCTCTG 1111
Qy 2492 lAGLYProHisHisAlaTrpAspGluGluProLYsProLeuLeuCysSerGlnTYrGluT 2512
Db 1110 CTGGCCCCCAACACGCTGGGAGAGAGGCCAAGCACTGCTGCTGCTGCAGTACGAGA 1051
Qy 2512 hTLeuSerAspSerGlu 2517
Db 1050 CACTCTCCGACAGCGAG 1034

Search completed: November 11, 2005, 10:58:53
Job time : 4786 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_p2n model

Run on: November 11, 2005, 09:39:08 ; Search time 5611 Seconds
(without alignments)
3709.701 Million cell updates/sec

Title: US-09-522-753-5
Perfect score: 13215
Sequence: 1 MSGSTQLVAQWTRATEPRYP.....WDEPRKLCSQYETLSDSR 2517

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Delop 6.0 , Delext 7.0

Searched: 9794790 seqs, 4134903567 residues
Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame_p2n.model -DEV=x1p
-Q=/cgn2_1/USFT0.epool_p/US09522753/runat_02112005.173641_20852/app_query.fasta.1.2695
-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MIMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS-human40.csi -LIST=45 -DOCCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pico -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09522753 @CGN 1.1 1531 @runat_02112005.173641_20852
-NCPUL=6 -ICPU=3 -NO_MMAP -LARGESUBSET -NEG_SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
- 20: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
- 21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
- 22: /cgn2_6/ptodata/1/pubpna/US10J_PUBCOMB.seq:*
- 23: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 24: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
- 25: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:*
- 26: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
- 27: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 28: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13215	100.0	7554	18 US-10-414-692-14	Sequence 14, Appl
2	13215	100.0	8561	18 US-10-174-014-11	Sequence 11, Appl
3	13215	100.0	8561	26 US-11-052-344-11	Sequence 1, Appl
4	13145	99.5	8657	14 US-10-087-192-653	Sequence 653, App
5	12978	98.2	7521	10 US-09-819-104A-3	Sequence 3, Appl
6	12978	98.2	8666	10 US-09-819-104A-1	Sequence 1, Appl
7	12978	98.2	8666	18 US-10-174-014-4	Sequence 4, Appl
8	12978	98.2	8666	21 US-10-723-860-1739	Sequence 1739, Ap
9	12845.5	97.2	9079	21 US-10-723-860-6114	Sequence 6114, Ap
10	10832.5	82.0	7386	10 US-09-819-104A-6	Sequence 6, Appl
11	10832.5	82.0	8544	10 US-09-819-104A-4	Sequence 4, Appl
12	10832.5	82.0	8544	22 US-10-764-420-1526	Sequence 1526, Ap
13	9691.5	73.3	7534	14 US-10-087-192-650	Sequence 650, App
14	4612	34.9	2930	16 US-10-146-473-14	Sequence 14, Appl
15	4612	34.9	2930	18 US-10-174-014-13	Sequence 13, Appl
16	4165.5	31.5	7940	18 US-10-341-434-235	Sequence 235, App
17	4165.5	31.5	7940	18 US-10-887-553A-664	Sequence 664, App
18	4147.5	31.4	7940	21 US-10-473-974-191	Sequence 191, App
19	4147.5	31.4	7940	22 US-10-956-157-1580	Sequence 1580, Ap
20	4147.5	31.4	7940	22 US-10-745-237-221	Sequence 221, App
21	3649	27.6	91141	14 US-10-087-192-649	Sequence 649, App
22	2951	22.3	23380	14 US-10-087-192-652	Sequence 652, App
23	2184	16.5	2745	18 US-10-414-692-15	Sequence 15, Appl
24	2133.5	16.1	22100	18 US-10-174-014-12	Sequence 12, Appl
25	1083	8.2	10910	24 US-10-745-237-23	Sequence 23, Appl
26	1083	8.2	10910	26 US-11-097-143-1937	Sequence 1937, Ap
27	1076	8.1	872	24 US-10-779-543-8757	Sequence 8757, Ap
28	880	6.7	718	22 US-10-278-698-515	Sequence 515, App
29	880	6.7	718	22 US-10-278-698-1031	Sequence 1031, Ap
30	746.5	5.6	520	9 US-09-920-300A-485	Sequence 485, App
31	746.5	5.6	520	14 US-10-033-528-485	Sequence 485, App
32	746.5	5.6	520	17 US-10-099-926-485	Sequence 485, App
33	746.5	5.6	520	24 US-10-961-527-485	Sequence 485, App
34	730.5	5.5	29467	22 US-10-937-730A-1	Sequence 1, Appl
35	730.5	5.5	67167	22 US-10-937-730A-3	Sequence 3, Appl
36	730	5.5	902568	16 US-10-156-761-1	Sequence 1, Appl
37	721	5.5	88400	22 US-10-844-716-1	Sequence 1, Appl
38	716.5	5.4	27705	10 US-09-942-025-12	Sequence 12, Appl
39	716.5	5.4	67311	10 US-09-942-025-11	Sequence 11, Appl
40	700	5.3	13416	16 US-10-205-032-11	Sequence 11, Appl
41	700	5.3	60196	16 US-10-205-032-11	Sequence 1, Appl
42	689.5	5.2	64492	19 US-10-378-083-1	Sequence 1, Appl
43	689	5.2	427	24 US-10-779-543-1895	Sequence 12895, A
44	685	5.2	30690	20 US-10-204-862A-1	Sequence 1, Appl
45	685	5.2	30690	26 US-11-005-196-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-10-414-692-14
; Sequence 14, Application US/10414692
; Publication No. US20030228607A1
; GENERAL INFORMATION:
; APPLICANT: X-Cepor Therapeutics, Inc.
; TITLE OF INVENTION: Screening method and modulators having an improved therapeutic
; FILE REFERENCE: 8012-002-US
; CURRENT APPLICATION NUMBER: US/10/414,692
; PRIOR FILING DATE: 2003-04-14
; PRIOR FILING DATE: 2002-04-15
; NUMBER OF SEQ ID NOS: 86

SOFTWARE: PatentIn version 3.2
SEQ ID NO 14
LENGTH: 7554
TYPE: DNA
ORGANISM: Homo sapiens
US-10-414-692-14

Alignment Scores:

Pred. No.:	0	Length:	7554
Score:	13215.00	Matches:	2517
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	18	Gaps:	0

US-09-522-753-5 (1-2517) x US-10-414-692-14 (1-7554)

QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTPdArgAlaThrGluProArgTyrPro 20
DB 1 ATGTGGGGCTCCACACAGCTGTGGCAGACGTGAGGGCCACTGAGCCCCCTTACCCG 60
QY 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
DB 61 CCCACAGCCTTCTCTACCAAGTGCAGATCGCCGAGCCACAGGACGTCGGGCTCTCG 120
QY 41 GlnTyrGlnHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
DB 121 GAGTACCAACACCACTCCGCGCATGTGCTCCCACTGTGCGCGGCTCCATCATCCAG 180
QY 61 ProGlnArgArgProSerLeuLeuSerGlnPheGlnProGlyValSngIuArgSerGln 80
DB 181 CCCACGGCGGAGGGCCCTCCCTGCTGTGAGTTCCACCCGGGATGAACGGTCCAG 240
QY 81 GlnLeuHisLeuArgProGlnSerHisSerTyrLeuProGlyLeuGlyLysSerGlnMet 100
DB 241 GAGCTTCACCTCGGCGCAAGTCCCACTATACCTCCGAGCTGGGGAAGTCAGAGATG 300
QY 101 GlnPheIleGlnSerLysArgProArgLeuGlnLeuLeuProAspProLeuLeuArgPro 120
DB 301 GAGTTCATTGAAAGCAAGCGCCCTCGGCTAAGAGCTCTGCTGCAACCCCTGCTGCAACG 360
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGlnAspLeuThrLysAspArgSer 140
DB 361 TCACCCCTCTGCGCCACGGGCGAGCTGTGGGATCTGAAGACCTCAACAGGACCGTTAC 420
QY 141 LeuThrGlyLysLeuGlnProValSerProProSerProProHisThrAspProGlnLeu 160
DB 421 CTGACGGGCAAGCTGGAACCGGTGTCTCCCGCCAGCCCCCGCACTGACCTTGAGCTG 480
QY 161 GlnLeuValProProArgLeuSerLysGlnLeuIleGlnAsnMetAspArgValAsp 180
DB 481 GAGCTGTGTCGCCGACAGGCTGTCCAAAGAGAGCTGATCCAGAACATGGAACCGCGTGA 540
QY 181 ArgGlnIleThrMetValGlnGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeu 200
DB 541 CAGAGATCAACATGATGAGCAGCAGATCTTAACCTCAAGAAAGAGAGCAGCAACGCTG 600
QY 201 GlnGlnGlnAlaAlaLysProProGlnProGlnLysProValSerProProIleGln 220
DB 601 GAGGAGAGAGGCTGCCAAGCCGCCGAGCTGAGAACCCGCTGTACCGCCGCTCATCGAG 660
QY 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGlnAsnArgLysLysAlaGlnAla 240
DB 661 TCGAAGCACCGCAGCTGTGGAGATCATCTAGCAGAGAACCGGAAGAGGCTGAAGCT 720
QY 241 AlaHisArgIleLeuGlnGlnLysGlnProGlnValGlnLeuProLeuTyrArgGlnPro 260
DB 721 GGACATCGGATTCGTGAAGAGCTGTGGGCGCCAGAGTGAAGCTGCGCTGTGAACAACG 780
QY 261 SerAspThrArgGlnTyrHisGlnAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
DB 781 TCCGACACCCGGGAGTATCATGAGAACATCAAAATTAACAGGCGATGCGGAAGAGCTA 840

QY 281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnThrTPdLysGlnLysPheCysGln 300
DB 841 ATCTTGTACTTCAGAGAGGAGAAATCACGCTCCGAAACATAGAGAGAAAGTTCTGCGCAG 900
QY 301 ArgTyrAspGlnLeuMetGlnAlaLeuGlnLysLysValGlnArgIleGlnAsnAsnPro 320
DB 901 CCTATGACCACTGAGTGAAGGCTTGGAAAAAGGTGAGCGCATGCAAAAACACCCG 960
QY 321 ArgArgArgAlaLysGlnSerLysValArgGlnTyrTyrGlnLysGlnPheProGlnIle 340
DB 961 CCCCCGGCGCCAGAGAGAGCAAGGTGGCGAGTACTAGAAAAGAGAGTTCCCTGAGATC 1020
QY 341 ArgLysGlnArgGlnLeuGlnGlnLysMetGlnSerArgValGlyGlnArgLysSerGly 360
DB 1021 CCGAAGCAGCGGAGCTGACGAGGAGCGCATGACAGAGGCTGGCGCAGCGGCGAGTGGG 1080
QY 361 LeuSerMetSerAlaAlaArgSerGlnHisGlnValSerGlnIleIleAspGlyLeuSer 380
DB 1081 CTGTCCATGTGGGCGGCCCGCAGCCAGACAGAGGTGCAGAGATCATCGATGGCTCTCA 1140
QY 381 GlnGlnGlnLysAsnLeuGlnLysGlnMetArgGlnLeuAlaValIleProPheMetLeuTyr 400
DB 1141 GAGCAGAGAGAACCTGGAAGAGCAGATGGCGCAGCTGGCCGTGATCCCGCCATGCTGTAC 1200
QY 401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspPheMet 420
DB 1201 GAGCTGACCCAGCAGCGCATCAAGTTCAATCAATGAAGGGCTTAATGGCCACCCCATG 1260
QY 421 LysValIleTyrLysAspArgGlnValMetAsnMetTPSerGlnGlnGlnLysGlnThrPhe 440
DB 1261 AAGGTGTACAAAGACCGCCAGGTCTATGAACATGTGAGTGAACAGAGAAAGAGACTTTC 1320
QY 441 ArgGlnLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGlnArg 460
DB 1321 CGGAGGAAGTTATCAGACATCCCAAGAACTTGGCGTATGTCATCTCCGAGAGAG 1380
QY 461 LysThrValAlaGlnCysValLeuTyrTyrTyrLeuThrLysLysAsnGlnLysTyrLys 480
DB 1381 AAGACAGGTGCTGAGTGTGCTCTTAATTAATCACTGACCTAAAGAAATAGAACTATAAG 1440
QY 481 SerLeuValArgArgSerTyrArgArgArgLysSerGlnGlnGlnGlnGlnGln 500
DB 1441 AGCTGTGTAGAGCGAGACTATGTGGCGCCGCGGCAAGCCAGCAGCAGCAACAGCAGCAG 1500
QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
DB 1501 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1560
QY 521 AspGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 540
DB 1561 GATGAGAAAGAGAAAGAAAGAGAGCGGAGAGAGAGAGAAACCCGAGGTGGAGAAC 1620
QY 541 AspLysGlnAspLeuLeuLysGlnLysThrAspAspThrSerGlyGlnLysAsnAspGln 560
DB 1621 GACAAAGAGAGACTCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
QY 561 LysGlnAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnIleYArgArgLysGly 580
DB 1681 AAGAGAGCTGTGCGCTCCAAAGCGCGCAAAATCTGCCAACAGCAGGAGAGAGCAAAAGGC 1740
QY 581 ArgIleThrArgSerMetAlaAsnGlnAlaAsnSerGlnGlnAlaIleThrProGlnGln 600
DB 1741 CCGATCAACCCGCTCAATGCTTAATGAGGCAACAGGAGAGAGCCATTCACCCCGCAGAG 1800
QY 601 SerAlaGlnLeuAlaSerMetGlnLeuAsnGlnLysSerArgTyrThrGlnGlnGlnMet 620
DB 1801 AAGGCGAGCTGTGCTCTCAATGAGCTGAATGAGATTTCTGCTGACACAAAGAGAAATG 1860
QY 621 GlnThrAlaLysLysGlyLeuLeuGlnHisGlyArgAsnTyrSerAlaIleAlaArgMet 640
DB 1861 GAAACAGCCAAAGAAAGTCTCTTGGAACAACGCGCAATGTGCGGCGCATCGCCCGAGATG 1920
QY 641 ValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGln 660

1921 GTGGGCTCCAAAGCTGTGTCCAGGTGAAGAACTTCACTTCAACTCAACAAAGAGCGAG 1980
Qy AsnLeuAspGluIleLeuGlnGlnHisIleValSerLeuMetGluValArgAsnAlaArg 680
Db AACCTCATGTAGTGACTTGTACAGACGACAAAGCTTGAAGATGAGAGAGAGAGAGAGCGCGG 2040
Qy ArgIleValSerValAspAlaProAlaAlaIleSerGluGluAlaAlaPheProProValVal 700
Db AGGAAGAAAGAGAAAGCGCGCGCGCGCGCGCGAGCGAGGCTGCACTTCCGCGCGTGGTG 2100
Qy GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGlu 720
Db GAGGATAGAGAGAGAGCGCTCGCGCTGAGCGGAATGAGAGAGAGATGTGAGAGAG 2160
Qy AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGlyCysSerGlyProAla 740
Db GCTAAAGCTTAACTGCTCTGGGAATGAGGTGCCAGAGGGGGAATGCAGTGGCCAGCC 2220
Qy ThrValAsnAsnSerSerAspThrGlySerIleProSerProHisIleThrGluAlaAlaVal 760
Db ACTGTCAACAAACAGCTCAGACACCGAGAGCATCCCTCTCTCAACAGTGAAGCGCGCAG 2280
Qy AspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProPro 780
Db GACACAGCGGAGAAATGGGCCCAAGCCCCAGCCACCTGGGCGCGGAGCGGCGCACCCCA 2340
Qy GluProProThrProProArgArgThrSerArgAlaProIleGluProThrProAlaSer 800
Db GGCCACCCACCCACCAACGAGAGACATCCGGGCGCCCATTTGAGCCACCCCGGCTCT 2340
Qy GluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProProPro 820
Db GAAGCCACCGGAGGCCCTTACGCCGCCCAACAGCACCCCATGCGCTTGCACCTCTCTCT 2400
Qy ValValProLysGluGluValGluGluThrAlaAlaProProValGluGluGly 840
Db GTGTCTCCCAAGAGAGAGAGAGAGAGAGACCCAGCACAGCGCCCAATGAGAGAGGG 2460
Qy GluGluGlnIysProProAlaAlaGluGluLeuAlaValAspThrGlyValAspGluGlu 860
Db GAGGAGCAGAAAGCCCCCGCGCGCTGAGAGAGCTGGCACTGACACAGGAGAGCGGAGAG 2520
Qy ProValIleSerGluCysThrGluGluAlaGluGluGlyProAlaValSerGlyLysAspAla 880
Db CCGGTCAAGAGCGAGTGCACGAGAGAGCCGAGAGAGGGCGCGCAAGGCGCAGAGAGCGG 2580
Qy GluAlaAlaGluAlaThrAlaGluGlyValAlaLeuValAspGluValSerGlyGlySer 900
Db GAGGCGCGCTGAGCGCACGGCGCGAGGGCGCGCTCAAGGACAGAGAGAGAGAGGGGAGC 2640
Qy GluAlaThrThrAlaValSerSerGlyValaProGluIleAspSerAspSerSerAlaThr 920
Db GGCAAGGCGCACCACTGCCAAGAGCTCGGCGCGCCCCAGGACAGCGACTCAGGTCAACC 2700
Qy CysSerAlaAspGluValaAspGluAlaGluGlyGlyValAspLysAsnArgLeuLeuSerPro 940
Db TGCAGTGCACACAGAGTGAAGTGAAGCGCGAGCGCGCGCAGAGAACCGGCTGTCTCCCA 2760
Qy ArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGluLysPro 960
Db AGGCGCGCGCTCTCTCAAGCTGCGAGACCCCGGCGCGCAATGCTCAACCCAGAAAGCCA 2820
Qy LeuAspLeuValGlnLeuValSerGluValaAlaAlaIleProProIleGluValThrLys 980
Db CTGGAACCTGAAGGAGCTGAAGAGCGAGCGAGCTGCATCCCGCCCATCCAGGTCAACAAA 2940
Qy ValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaProPro 1000
Db GTCCATGAGCGCCCCCGGAGAGAGCGAGCTCCACCAAGCAGAGTCTCCCGAGCCCGCAG 3000
Qy ProProGluAsnLeuGluIleProGluSerAspAlaProGluGlnProGlySerSerProArg 1020

3001 CCACCGAAMAACCTGACCGCGAGAGCGAGCGCCCTCAGGACGCTTGGCAGCAGCGCCCG 3060
Qy GlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGln 1040
Db GGCAGAGACAGAGACCCCGGACCCCCCGCGCACAAAGAGGCTTTCGAGCGAGGCGCAG 3120
Qy LysLeuProGlyAspProProCysIleProThrSerGlyLeuProPheProValProProArg 1060
Db AAGCTGCTGGGAGACCCCGCTTGTGAGCTTCCGGCGCTGCCCTTCCCGTGCCTCCCGCT 3180
Qy GluValIleIleValAlaSerProHisAlaProAspProSerAlaPheSerTyAlaProPro 1080
Db GAGGTGATCAAGGGCTCCCGCATAGCCCGGAGCCCTCAGGCTTCTCTACGCTTCAACT 3240
Qy GlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProPro 1100
Db GGTACCCACTGCCCCCTGGGCTTCAAGACATGCGCGGCGCGTCTCGCGCGCGCACCC 3300
Qy ThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeuGluArg 1120
Db ACCATCTCCAACCGCGCTCCCTCATCTCTGTGCAAGACCCCAAGCTCTCGAGAGG 3360
Qy GlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTySerGluHis 1140
Db CAATAGTGCATCTCCCAAGGAATGCTCGTCCAGCTCCACGCTCCGTTACTCAGAGCAT 3420
Qy AlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLys 1160
Db GCCAAGGCGCGGGGCGCTGTACATAGGGGTGCGCTTGCCTCAATGAGACCCCAAAAG 3480
Qy LeuAlaProPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyPro 1180
Db CTGGACACCTTCAAGCGAGAGAGAGAGAGACAGTGTCCCAAGGGCGCAGGCTGGGCA 3540
Qy ProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeu 1200
Db CCGGAGAGCTTGGGGGTGCCACAGGCCAAGAGGGCGTCCGTGTGAGAGGAGACAGCTTG 3600
Qy GlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAsp 1220
Db GGCTCAGTTCCGGAGCGAGACATACCAAGAGCATTTCCAGCACAGCGGTGCTCCGAGC 3660
Qy SerAlaIleThrTyArgGlySerIleThrHisGlyThrProAlaAspValLeuTyLys 1240
Db AGCGCATCATCAATCGCGGCTCATCAACCCAGGAGCGCACGTGCTGTACAG 3720
Qy GlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGlu 1260
Db GGCAACATCACAGATCATCGCGAGGAGACGCCGAGTGTGGACCGCGCGCGGAG 3780
Qy AspSerLeuProLysGlyHisValIleTyArgGluGlyValLysGlyHisValLeuSerTy 1280
Db GACAGCTGCGCAAGGCGCACGTATCTAGAAAGCAAGAGGCCACGCTTGTCTAT 3840
Qy GluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyPro 1300
Db GAGGTGCGCATGTCTGTGACCCAGTGTCCAAAGAGAGCGCAGAGAGCTCAGAGCCC 3900
Qy ProHisGluThrAlaAlaProLysArgThrTyAspMetMetGluGlyArgValGlyArg 1320
Db CCCCATGAGACGGCGCGCCCAAGCGCACCTTAGATGATGAGGGCGCGGTGGGAGCA 3960
Qy AlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHis 1340
Db GCCATCTCTCAAGCCAGCATGAAAGTCTCATGGCGGTGCATCCCGCGGAGCGACAC 4020
Qy SerProHisIleLeuLysGluGlnHisAlaArgGlySerIleThrGlnGlyIlePro 1360
Db AGCGCCACCACTTCAAGAGAGAGAGACACATCCGGGGTCCATTCACAAAGGAGATCCCT 4080
Qy ArgSerTyValGluAlaGlnGluAspTyLeuArgArgGluAlaLysLeuLeuValArg 1380
Db CGGTCTAGTGAAGCAGAGAGAGCTACCTGCGTGGGAGGCGCAAGCTCTTAAAGCGG 4140

QY 1381 GIUGIYThrProProProProProSerArgAspLeuThrGluAlaTyrIleThrGln 1400
DB 4141 GAGGGAGAGCCTCCGCGCCACCGGCTCACGGGACCTGACCGAGGCTTAACAAGACGAG 4200
QY 1401 AlaleuGlyProLeuLeuLeuLeuProAlaHisGluGlyLeuValAlaThrValIlySerGlu 1420
DB 4201 GCCCTGGGGCCCCCTGAAGCTGAAGCCGGCCCATGAAGGGCTGGTGGCCACGGGTGAAGGAG 4260
QY 1421 AlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuPro 1440
DB 4261 GGGGGCCGCTCCATCATAGATCCGCGCGAGAGACTGGGCAACGCCCGAGCTGCCCC 4320
QY 1441 LeuAlaProArgProLeuLeuGlySerIleThrGlnGlyThrProLeuLeuTyrAsp 1460
DB 4321 CTGGCCCCCGGCGGCTCAAGAGGGGCTCCATCACGAGGGACCCCGCTCAAGTACGAC 4380
QY 1461 ThrGlyAlaSerThrThrGlySerIlySerIlyHisAspValArgSerLeuIleGlySerPro 1480
DB 4381 ACCGGGCGCTCCACCTGGCTCCAAAAGACGAGTACGCTCCCTCATCGGAGCCCC 4440
QY 1481 GlyArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGln 1500
DB 4441 GGCCGGAGCTTCCACCCCTGCACCCGCTGGATGTATGGCCGACCCCGGGGCACTGGAA 4500
QY 1501 ArgAlaCysTyrGluGluSerLeuLeuSerArgProGlyThrAlaSerSerGlyGly 1520
DB 4501 CGTGCTCTGCTACGAGGAGAGCTGAAGAGCCGGCCAGGACCGCCAGCAGCTCGGGGGC 4560
QY 1521 SerIleAlaArgGlyAlaProValIleValProGluLeuGlyIlyProArgGlnSerPro 1540
DB 4561 TCCATTGGCCGGCGGCCCCCGGCTATTGTGCTGACCTGGTAAAGCCGGCGAGCCCC 4620
QY 1541 LeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisGluProArgGlySerProVal 1560
DB 4621 CTGACCTATGAGGACCAACGGGGGCACTTTGCCGGGCACTCCCAAGAGTTGCCCGG 4680
QY 1561 ThrMetArgGluProThrProArgLeuGlnGlySerLeuSerSerSerIlyAlaSer 1580
DB 4681 ACCATCGGGAGCCACCGCGGCTGACGAGAGGAGCACTTTGTCGACAAAGGATCC 4740
QY 1581 GlnAspArgIlyLeuThrSerThrProArgGluIleAlaIlySerProHisSerThrVal 1600
DB 4741 CAGGACCGGAAGCTGACGTGAGCCTCGTGAATGCGCAAGTCCCGGACAGCACCGG 4800
QY 1601 ProGluHisHisProHisProIleSerProTyrGluHisGluLeuArgGlyAlaSerGly 1620
DB 4801 CCGGAGCACCAACCAACCCCATCTGCTGATGAGCACTGCTTGGGGGCGTGAATGGC 4860
QY 1621 ValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGly 1640
DB 4861 GTGGACCTGATCGACGACATCCCTGGCTTGACCCCACTCCATACCCCGCGGC 4920
QY 1641 IleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisGluAlaProAspProThr 1660
DB 4921 ATCCCTCTGAGACGACCGCTGCTTACTACCTGCGCCGACACTGGCCCCCAACCCAC 4980
QY 1661 TyrProHisGluTyrProProTyrLeuLeuIleArgGlyTyrProAspThrAlaAlaLeuGln 1680
DB 4981 TACCCGCACTGTACCACTTACTCTACTCCGGGCTTACCCGACACGGCGGGGCTGGAG 5040
QY 1681 AsnArgGlnThrIleIleLeuAspTyrIleThrSerGlnGlnMetHisHisAsnThrAla 1700
DB 5041 AACCGGACACCATCATCATGACTACATCACTCCGACAGATGACCAACAACAGCGGC 5100
QY 1701 ThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeu 1720
DB 5101 ACCGCGCATGGCCGAGGAGCTGATATGCTGAAGGGGCTCTCGCCCGGCGAGTCTCTGCTG 5160
QY 1721 AlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisGlu 1740
DB 5161 GCACTAACTTACGCTCGGGGTCCCGAGGACATGACCTGTGCCAAGTGCACACCTG 5220

QY 1741 ProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyr 1760
DB 5221 CCTGTGCTGTGTCGCCCGACACAGGACCCGAGCACCGCATGAGCGCTTGCTTAC 5280
QY 1761 LeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGly 1780
DB 5281 CTCCCCACCGCGCCCCAGCCCTTACAGACCGCCCAAGAGCTCCCACTCTCCAGGA 5340
QY 1781 GlyProThrHisGluThrIlySerProThrThrThrSerSerSerGluArgGluArgAspArg 1800
DB 5341 GGTCCAAACACTTGACAAAACCAACACAGTCTCTGCTCGAGGGGAGGAGCACCGG 5400
QY 1801 AspArgGluArgAspArgAspArgGluArgIlySerSerIleLeuThrSerThrThrThr 1820
DB 5401 GATCAGAGCGGGAGCCGGATGGAGCGGAAAGTCCATCTCACTCCACACAGAG 5460
QY 1821 ValGluHisAlaProIleTyrArgProGlyThrGlnGlnSerSerGlySerSerGlySer 1840
DB 5461 GTGAGCACGACCCATCTCGAGACTGTGACAGCTGTACAGACAGCAGCGGCGAGC 5520
QY 1841 SerGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHis 1860
DB 5521 AGCGCGGGGGTGGGGGAGAGAGCGAGCGCGCCGCTCCCATCCCATCCACAGCAG 5580
QY 1861 SerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsn 1880
DB 5581 TCGCCCATCTCTCCCTGACCCGAGATGCTTCACAGACAGACCCAGTGTGCTTCAAC 5640
QY 1881 ThrGlyMetIlyGlyIleIleThrAlaValGluProSerIlyProThrValLeuArgSer 1900
DB 5641 ACGGCACTAAGGCTATCATCACCGCTGTGAGCCCAAGAACCCAGCGTCTCGAGGTCC 5700
QY 1901 ThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCysPro 1920
DB 5701 ACCTCACTCTCTCAACCGCTGCGCAGCTGCAATTCCTCCACTGCAACCCACCTGCCA 5760
QY 1921 LeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGlnProValLeuLeuProIlys 1940
DB 5761 CTGGGGGGGACCTCGATGGGGGTCTACCTTACCTCATGAGCCCGCTTGTGCTGCAAG 5820
QY 1941 GluAlaProArgValAlaArgProGluArgProArgProAlaAspThrGlyHisAlaPheLeu 1960
DB 5821 GAGGCCCCCGGGGTGCCCGGCAAGCGGCCCCGACAGACACCGGCATGCTTCTTC 5880
QY 1961 AlaIlyProProAlaArgSerGlyLeuGluProAlaSerSerProSerIlySerGly 1980
DB 5881 GCCAAGCCCCAGCCGCTCGGGCTGAGCCCGCTCTCCCGCCAGCAAGGCTCGGAG 5940
QY 1981 ProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaIlys 2000
DB 5941 CCGCGCCCTTGTAGTGTCTCTGTGTGCGCACGGCCACCATGCGCACCCCTGGAAAG 6000
QY 2001 AsnLeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSerAsp 2020
DB 6001 AACCTGCACTCACCAACCGGACCGCGGCGGCCACATGCTCGGCGCTCGGAC 6060
QY 2021 ProHisArgGluIlyThrGlnSerIlySerProPheSerIleGlnGluLeuLeuArgSer 2040
DB 6061 CCGCACCGGGAAGAAGCTCAAGTAAACCTTTTCATGCAAGAACTGGAACTCCCTTCT 6120
QY 2041 LeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSer 2060
DB 6121 CTGGGTTCACAGGAGGAGCTTACAGCCCCCGAAGGGGTGAGACCCCTGAGCCCTTGAGC 6180
QY 2061 SerProSerLeuThrHisAspIlySerGlyLeuProIlyHisGluGluLeuAspIlySer 2080
DB 6181 TACCCAGTCTGACCCACAGCAAGGGGCTTCCCAAGCACTGGAAGAGCTCCAGCAAGAGC 6240
QY 2081 HisLeuGluGlyGluLeuArgProIlySerGlnProGlyProValIlyLeuGlyGlyAla 2100
DB 6241 CACTGGAAGGGAGGAGCTGGGCGCAAGAGCAGGCGCCGTGAAGCTTGGCGGGAGGCGC 6300
QY 2101 AlaHisGluProHisGluArgProLeuProGluSerGlnProSerSerSerProLeuLeu 2120


```

Db      6301 GCCACCTCCACACCTGGCGGCGCTGCTGAGAGCCAGCCCTGCTCAGCCCTGCTC
Qy      2121 GlnThrAlaProGlyValIleGlyGlyHisGlnArgValValThrLeuAlaGlnHisIleIleSer
Db      6361 CAGACCCGCCCCAGGGGCTCAAGGTCACAGCGGGGGTGGTACCCCTGGCCCAACATCATGCT
Qy      2141 GluValIleThrGlnAspTyrThrArgHisHisAspProGlnGlnLeuSerAlaProLeuPro
Db      6421 GAGGTATCAACAGAGCTACACCCGGGACACCCAGAGCTCAAGGCACTCCCTGCCC
Qy      2161 AlaProLeuTyrSerPheProGlyAlaSerCyAspProValLeuAspLeuArgArgProPro
Db      6481 GCCCCCTCTACTCTCTCTCCCTGGGGCAGCTGCCCCCTCTGAGACTCCGCCGCCACCC
Qy      2181 SerAspLeuTyrLeuProProProAspHisGlyValAlaProAlaArgGlySerProHisSer
Db      6541 AGTGACCTTACCTCCGCGCCCGGACCAATGATGTCCTCCGCGCTGCTCCGCCACAGC
Qy      2201 GluGlyGlyIleAspArgSerProGluProAsnIleThrSerValLeuGlyGlyGlyIleAsp
Db      6601 GAAAGGGGCAAGAGTCTCCAGAGCCAAACAGACGTGCTGGTGGTGGTGGTGGAGAC
Qy      2221 GlyTlIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAla
Db      6661 GGTATTGAACTGTGTCCCAACCGAGGATGACGAGGCACTCCCGGAGTGCT
Qy      2241 ValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerIle
Db      6721 GTGTACCCGCTGCTGTAACCGAGATGGGAAACAGCGGAGCCAGCAAGATGGGCTTCAG
Qy      2261 SerProGlyAsnThrSerGlnProProAlaPhePheSerIleThrGluSerAsnSer
Db      6781 TCTCCAGGCAACACAGCCAGCCCGCCGCTTCTTCAGCAAGCTGACCGAGAGCACTCC
Qy      2281 AlaMetValIleSerIleValGlnGluIleAsnIleValLeuAsnThrHisAsnArgAsn
Db      6841 GCCATGTCAGTCAAGTCAAGAGCAAGAGATCAACAGAGCTGAAACCAACCCGAGAT
Qy      2301 GluProGluTyrTrpMetIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThr
Db      6901 GAGCCCTGAATACATATACAGCCAGCTGAGCGAGATCTTCAATATAGCCCGCATACCC
Qy      2321 GlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGlnIleHisAlaSerThrAsnMet
Db      6961 GGAACAGGCTTATGACTTATAGAGCCAGCGGTGAGAAACATGCCACACCAACATG
Qy      2341 GlyLeuGluAlaIleIleArgIleValIleLeuMetGlyIleTyrArgProGlnTyrGluSer
Db      7021 GGGCTGAGGCGCATTAATTAGAAAGGCACTATGGGTAAATATAGCACTGGGAAAGATCC
Qy      2361 ProProLeuSerIleAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaIle
Db      7081 CCGCCCTCAGCGCCCAATGCTTTTAAACCTCTGAATCCAGTCCAGCCCTGCGCTGCT
Qy      2381 MetProIleThrAlaIleAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGly
Db      7141 ATGCCCATTAACCTGCTGACGAGCGAGTGCACACACTTCACTTGCACAGGTGGGAGC
Qy      2401 GlyIleAlaIleValIleSerGlyArgProSerSerArgIleValIleSerProAlaProGly
Db      7201 GGGAAAGGCCAAGGCTCTGGCAGACCCAGACCGCAAAAGCCAAAGTCCCGGCGCGGCG
Qy      2421 LeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluGlyAspCyAsn
Db      7261 CTGGCACTCTGGGAGCGGCACTCTGTCTCTCAAGTGCATCGAGGAGACTCCAAC
Qy      2441 ArgArgThrProLeuThrAsnArgValTyrGluAspArgProSerSerAlaGlySerThr
Db      7321 CGCCGAGACGCGCTCAACACCGCGTGTGGAGAGACAGCGCTCTGTCGAGGTTTCAGC
Qy      2461 ProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyAlaMetAlaSerProPro

```

```

Db      7381 CCATTCCTCAACACCCCTGATCATGCGGTGAGGCGGGTGTATGGCTTCCCAACC
Qy      2481 ProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaTyrAspGlu
Db      7441 CCACCGGCGCTCCCGCGGCGAGCGGCGCCCTGCTGCTGCGCCCAACAGCTGGGAGCAG
Qy      2501 GluProIleProLeuLeuCySerGlnTyrGluThrLeuSerAspSerGlu
Db      7501 GAGCCCAAGCACTGCTCTGCTGCTGCACTACGAGCACTTCCGACAGCAG
RESULT 2
US-10-174-014-11
; Sequence 11, Application US/10174014
; Publication No. US20040005292A1
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Susan M. Freiler
; APPLICANT: Kenneth W. Dobie
; TITLE OF INVENTION: ANTISENSE MODULATION OF SMRT EXPRESSION
; FILE REFERENCE: PTS-0012
; CURRENT APPLICATION NUMBER: US/10/174,014
; CURRENT FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 73
; SEQ ID NO 11
; LENGTH: 8561
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2) ... (7555)
US-10-174-014-11
Alignment Scores:
Pred. No.: 0 Length: 8561
Score: 13215.00 Matches: 2517
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 18 Gaps: 0
US-09-522-753-5 (1-2517) x US-10-174-014-11 (1-8561)
Qy      1 MetSerGlySerThrGlnLeuValAlaGlnThrTyrArgAlaThrGluProArgTyrPro
Db      2 ATGTCCGGCTTCCACACAGCTTGTGSCACAGACTGAGGGCCACTGAGCCCTTACCCG
Qy      21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu
Db      62 CCCACAGCTTCTTCTTACCAAGTGCAGATGCGCCGAGCGACAGCGAGTGGGCTCTG
Qy      41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln
Db      122 GAGTTCACAGACACATCCCGCGCACTATGCTCTCCACCTGTGCGGGCTTCATATCCAG
Qy      61 ProGlnArgArgArgProSerLeuLeuSerGluPheGluProGlyValGlnIleValGlySerGln
Db      182 CCCACAGCGGAGGCGCTCTCTGCTGCTGAGTTCAGAGCCGAGAAATGACGATCCAG
Qy      81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyIleYlySerGluMet
Db      242 GAGCTCACCTGCGGCGCAGAGTCCCACTATACCTCCCGAGCTGGGAAAGTCAAGATG
Qy      101 GluPheIleGluSerIleAspArgProAlaGluGluLeuLeuProAspProLeuLeuArgPro
Db      302 GAGTTCATTGAAAGCAAGGCGCTCGCTAGAGCTGCTCTGAGACCCCTGCTGACACG
Qy      121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGlyAspLeuThrIleYlyAspArgSer
Db      362 TCACCCCTCTGCGCCACCGGCGCAGCTCGCGGATTCGAAAGACTTCACCAAGACCTTACG
Qy      141 LeuThrGlyIleLeuGluProValSerProProSerProProHisThrAspProGluLeu
Db      422 CTGACGCGGCAAGCTGGAACGGGTGTCTCCCGCAGCGCCCGCACACATGACCTTGAGCTG

```

QY 161 GIuleuValProProArgLeuSerLySGluLeuIleGlnAsnMetAspArgValAsp 180
Db 482 GAGCTGTGCCCCCAACGGCTGTCCAAAGAGAGCTCATCCAGAACATGAGACCGCGTGGAC 541
QY 181 ArgGluIleThrMetValGluGlnGlnIleSerLySLeuLySLeuLySGlnGlnIleu 200
Db 542 CAGAGATCATCATGTGTAGACAGACGATCTTAACCTCAAGAGAGAGACGAAACGCTG 601
QY 201 GluGluGlnAlaAlaLySProProGluProGluLySProValSerProProIleGlu 220
Db 602 GAGAGAGAGGCTGCGCAAGCCCGCCAGCTCGAGAGACCCGTGTCAACCGCCCATCGAG 661
QY 221 SerLySLeuArgSerLeuValGlnIleIleTyAspGluAsnArgLySLeuValaGlu 240
Db 662 TCGAAGCACCGCAGCGCTGTGAGATCATCTAGCAGAGAACCGGAAGAGGCTGAAAGCT 721
QY 241 AlaHisArgIleLeuGlnGluLySLeuGlyProGlnValGluLeuProLeuTyThrSngInPro 260
Db 722 GCACATCGGATCTGGAGAGGCTCGGGGCCCAAGTGAAGCTGCCCTGTACAAACGAGCC 781
QY 261 SerAspThrArgGluThrHisGluAsnIleLySLeuGlnAlaMetArgLySLeu 280
Db 782 TCCGACACCCGGCAGTATCATAGAACATCAAAATTAACCGCCATGCCGAAGAGCTA 841
QY 281 IleLeuTyThrPheLySArgArgAsnHisAlaArgLySGlnTrpLySGlnLySPhCySgIn 300
Db 842 ATCTTGTACTTCAAGAGAGGAAATCACGCTCGGAACAAATGGAAGCAAGATTCTCCAG 901
QY 301 ArgTyThrAspGlnLeuMetGluAlaLeuGlnLySLeuValGluArgIleGlnAsnAsnPro 320
Db 902 CCTATAGACACAGCTCATGAGGCGCTTGGAAGAAAAAGTGAAGCGCATCGAAAAACACCG 961
QY 321 ArgArgArgAlaLySLeuSerLySValArgGluTyThrGluLySGlnPheProGluIle 340
Db 962 CGCGCGCGGCGCAAGAGAGCAAGGTGCGAGTACTACGAAAAAGATTCCCTGAGATC 1021
QY 341 ArgLySGlnArgGluLeuGlnGlnIleuArgMetGlnSerArgValGlyGlnArgLySergLy 360
Db 1022 CCCAAGCACCGGAGCTGCAGAGAGCGCATGCAGACAGCGGTGGCCAGCGCGGCAATGGGG 1081
QY 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
Db 1082 CTGTCCATGTCCGGCCCCCGCAGCGAGCAAGAGTGTCAAGATCATCGATGCCCTCTCA 1141
QY 381 GluGlnGlnAsnLeuGlnLySLeuGlnMetArgGlnLeuAlaValIleProProMetLeuTy 400
Db 1142 GAGCAGGAGAACCTGGAGAGAGAGATGCGCACTGCGGTGATCCCGCCATGCTGTAC 1201
QY 401 AspAlaAspGlnGlnArgIleLySPhelIleAsnMetAsnGlyLeuMetAlaAspProMet 420
Db 1202 GACGCTGACACAGAGCGCATCAAGTTCAATCAATGACGGGCTTATGGCCGACCCCATG 1261
QY 421 LySValTyThrAspArgGlnValMetAsnMetTrpSerGluGlnGlnLySLeuTrpPhe 440
Db 1262 AAGGTGTAAACAAGACCGCAGAGTCATGAACATGTGAGTGAAGACGAGAGAGACCTTC 1321
QY 441 ArgGlnLySPhMetGlnHisProLySAsnPhelGluLeuIleAlaSerPheLeuGlnArg 460
Db 1322 CGGGAAGGTCATGACATCCAGAACCTTGGCGCTGATGCATATCTCGAGAGAG 1381
QY 461 LySThrValAlaGluCySValLeuTyThrTyThrLeuThrLySLeuAsnGluAsnTyLyS 480
Db 1382 AAGACAGTGGCTAGTGGCTCTCTATTACTACCTGACTAAGAAATGAGATGAACTTAAG 1441
QY 481 SerLeuValArgArgSerTyThrArgArgGlyLySLeuGlnGlnGlnGlnGln 500
Db 1442 AGCTGTGTGAGAGAGAGTATCGGCGCGCGCAAGGACGACGACGACAAACGACGACG 1501
QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
Db 1502 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1561

QY 521 AspGluLySGluLySGluLySGluAlaGluLySGluGluGluLySProGluValGluAsn 540
Db 1562 GATGAGAAAG 1621
QY 541 AspLySGluAspLeuLeuLySGluLySThrAspAspThrSerGlyGluAspAsnAspGlu 560
Db 1622 GACAG 1681
QY 561 LySGluAlaValAlaSerLySGlyValArgLySThrAlaAsnSerGlnGlyArgArgLyS 580
Db 1682 AAGAGAGGCTGTGCTCCAAAGCGCGCAAAATCTGCCAAGCAGAGAGAGAGAGAGAGAG 1741
QY 581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600
Db 1742 CCGATCACCCGCTCAATGCTATATAGGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1801
QY 601 SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluMet 620
Db 1802 AGCGCGAGCTGGCTCTCCATGGAGCTGAATGAGATTCTCGCTGACAGAGAGAGAGAGATG 1861
QY 621 GluThrAlaLySLeuGlyLeuLeuGlnHisGlyArgAsnTrpSerAlaIleAlaArgMet 640
Db 1862 GAAACAGCCAG 1921
QY 641 ValGlySerLySThrValSerGlnCySLeuAsnPhetyrPheAsnTyThrLySArgGln 660
Db 1922 GTGGGCTCCAGACGTGTGTGAGGTGAACCTTCACTTCACTCAAGAGAGAGAGAGAGAG 1981
QY 661 AsnLeuAspGluIleLeuGlnGlnHisLySLeuLySMetGluLySGluArgAsnAlaArg 680
Db 1982 AACCTCGATGAGATCTTGACAGACAGACAGAGCTGAAGATGAGAGAGAGAGAGAGAGAG 2041
QY 681 ArgLySLeuLySLeuAlaProAlaAlaAserGluGluAlaAlaPheProProVal 700
Db 2042 AAG 2101
QY 701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGlu 720
Db 2102 GAGAGTGAAGAGATGAGAGCGTCCGGCGTGAACCGGAAAGAGAGAGAGAGAGAGAGAG 2161
QY 721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCySergLyProAla 740
Db 2162 GCTGAAGCTTCAATGCTCTTGGAATGAGGTGCCAGAGGGAGATGAGGTGCCAGGCC 2221
QY 741 ThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLyS 760
Db 2222 ACTGTCAACAACAGCTCAACACCGAGAGCATCCCTCTCTCAACACTGAGGCGCCAGAG 2281
QY 761 AspThrGlyGlnAsnGlyProLySProProAlaThrLeuGlyValAspGlyProProPro 780
Db 2282 GACACAGGCGCAGAGATGGGCCCAAGCCCGCCAGCCACTTGCGCGCGCAGCGGCCACCCCA 2341
QY 781 GlyProProThrProProArgArgThrSerArgAlaProIleGluProThrProAlaSer 800
Db 2342 GGCCCAACCCACACAGAGAGAGAGATCCGGGCCCAATTGAGGCCACCCGGGCTCT 2401
QY 801 GluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProProPro 820
Db 2402 GAAGCCACCGAGGCCCTTACGCCCCACAGACACCCCATCGGCTCTGACCTCTCTCT 2461
QY 821 ValValProLySGluGluLySGluGluGluThrAlaAlaAlaProProValGluGluGly 840
Db 2462 GTGGTCCCAAG 2521
QY 841 GluGluGlnLySProProAlaAlaGluGluLeuAlaValAspThrGlyLySAlaGluGlu 860
Db 2522 GAGAGCAG 2581
QY 861 ProValLySergLySLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 880
Db 2582 CCGGTCAAG 2641
QY 881 GluAlaAlaGluAlaThrAlaGluGluGluAlaLeuLySLeuLySGluGlyGlySer 900

DB 2642 GAGGCCCTGAGGCGACGGCGGCGCTCAAGCAGAGAAAGAGGCGGAGC 2701
QY 901 G1YArg1aThrThrAlaYsSerSerG1YAlaProGlnApsSerApsSerSerAlaThr 920
DB 2702 GGCAGGGCCACCACTGCCAAGAGCTCGGGCGCCCCAGCAGCAGCACTCCAGTGTACC 2761
QY 921 CysSer1aApsG1YAlaApsG1YAlaG1YG1YAlaApsYsApsA1YLeuLeuSerPro 940
DB 2762 TGCAGTGCAGACGAGGTGATGAGCGCGAGGGCGGCAAGAACCGGTCTCTCTCCCA 2821
QY 941 ArgProSerLeuLeuThrProThrG1YAspProArg1aApsA1YAspProGlnLYsPro 960
DB 2822 AGGCCCAAGCTCTCCACCCGAGCTGGCGAGCCCCCGGCGCAATGCTCAGCCCAAGAACCA 2881
QY 961 LeuApsLeuYsGlnLeuYsGlnArgAlaAlaAla1LeuProPoliLeuAlaThrLYs 980
DB 2882 CTGGACCTGAAGCAGCTGAAGAGCAGCGGCTGCCATCCCCCATTCAGGTCAACAAA 2941
QY 981 ValH1eG1YProProArgG1YAspAlaAlaProThrYsProAlaProProAlaProPro 1000
DB 2942 GTCCATGAGCCCCCGGAGAGAGCAGCTCCCAAGCCAGTCCCCAGAGCCCAACG 3001
QY 1001 ProProGlnApsLeuGlnProG1YAspAlaProGlnGlnProG1YSerSerProArg 1020
DB 3002 CCAACCGCAAACTGACCGCGAGAGCAGCGCCCTCAGCAGCTCGGACAGAGCCCCGG 3061
QY 1021 G1YLYsSerArgSerProAlaProProAlaApsLYsG1YAlaApsAlaG1YAlaGln 1040
DB 3062 GGCAGAGAGCAGAGCCCGGAGCCCCCGCAGCAAGAGGCTTCGAGCGGAGGCCAG 3121
QY 1041 LYsLeuProG1YAspProProCySTPThrSerG1YLeuProPheProValProProArg 1060
DB 3122 AACCTGCTGGGAGCCCCCTTGCTGATCTTCGGGCTGCTCCCGTCCCGTCCCGCCGT 3181
QY 1061 G1YVal11eLYsAlaSerProH1aAlaProApsProSerAlaApsSerTYrAlaProPro 1080
DB 3182 GAGGTATCAAGAGCTCCCGGCACTGCGGAGCCCCGACCTTCCTCTAGCTCCACCT 3241
QY 1081 G1YH1aProLeuProLeuG1YLeuH1aApsThrAlaArgProValLeuProArgProPro 1100
DB 3242 GGTACCCCACTGGCCCTGGGCTCCATGACATGCGCCGCTGCGCGGCCCAACCC 3301
QY 1101 Thr11SerApsApsProProLeuH1eSerSerAlaYsH1aProSerVal1LeuG1YArg 1120
DB 3302 ACCATCTCCAAACCGGCTCCCTCATCTCTCTGCCAAGCAACCCAGCTCTCGAGAGG 3361
QY 1121 Gln11eG1YAla11eSerGlnG1YMetSerValGlnLeuH1eValProTYrSerG1YH1e 1140
DB 3362 CAATAGGTGCCATCTCCCAAGGAATGTCGATCAGCTCCAGCTCCCGTACTCAGAGCAT 3421
QY 1141 AlaLYsAlaProAlaG1YProVal1ThrMetG1YLeuProLeuProMetApsProLYs 1160
DB 3422 GCCAAGAGCCCGGTGGGCTCTGTCACCATGGGCTGCTGCCATGAGCCCAAAAG 3481
QY 1161 LeuAlaProPheSerG1YValLYsGlnG1YLeuSerProArgG1YGlnAlaG1YPro 1180
DB 3482 CTGGACCTTCACGCGAGAGAGAGAGAGCACTGTCCCAAGGGGCCAGGTGGGCCA 3541
QY 1181 ProG1YSerLeuG1YAlaProThralaGlnG1YAlaSerValLeuArgG1YThralaLeu 1200
DB 3542 CCGAGAGGCTGGGGGTGCCACAGCCAGAGAGGCTCCGTGCTGAGAGGAGCAGCTGTG 3601
QY 1201 G1YSerValProG1YG1YSer11eThrLYsG1Y11eProSerThrArgValProSerAps 1220
DB 3602 GGCTCACTTCGGGGCGAAGCATCACAAAGGCAATCCCAAGCAGAGGGGTGCTCGAGC 3661
QY 1221 SerAla11eThrTYrArgG1YSer11eThrH1eG1YThrProAlaApsValLeuTYrLYs 1240
DB 3662 AGGCGCATCAACATCGCGGCTCATCAACCAAGGAGCGCACTGACGTCTGTACAG 3721
QY 1241 G1YTh11eThrArg11e11eG1YG1YApsSerProSerArgLeuApsArg1YArgG1Y 1260

DB 3722 GGCACATCAACAAGATCATGTGGGAGAGACGCCGAGTGTGACCGCGCGGAG 3781
QY 1261 ApsSerLeuProLYsG1YH1eVal11eTYrGlnG1YLYsG1YH1eValLeuSerTYr 1280
DB 3782 GACAGCTGGCCCAAGGGCCAGGTATACAGAGGCAAGAGGGCCAGTCTTCTCAT 3841
QY 1281 GlnG1YLYsSerVal1ThrGlnCYsSerLYsG1YApsG1YArgSerSerSerG1YPro 1300
DB 3842 GAGGTGAGATGTCTGTGACCCAGGTCTCCAAGAGAGACCGCAGAACAGTCAAGACCC 3901
QY 1301 ProH1eG1YThralaAlaProLYsArgThrTYrApsMetMetGlnG1YArgValG1YArg 1320
DB 3902 CCCATGAGACGGCGCGCCCAAGCGCACTATGACATGATGAGGCGCGCTGTGGGAGA 3961
QY 1321 Ala11eSerSer1aSer11eG1YLeuMetG1YArgAla11eProProG1YArgH1e 1340
DB 3962 GCCATCTCTCAGCCAGATGAAAGTCTATGAGCGGTTCATCCCGCGAGCGACAC 4021
QY 1341 SerProH1aH1eLYsG1YGlnH1eH1eAlaArgG1YSer11eThrGlnG1YLePro 1360
DB 4022 AGCCCCACACCTCAAGAGAGCAGCAGCATCCGGGTCTCATCAACAAGGATCTCT 4081
QY 1361 ArgSerTYrValG1YAlaGlnG1YApsTYrLeuArgArgG1YAlaLYsLeuLeuLYsArg 1380
DB 4082 CGGTCTCTACGTGAGGACAGAGGACTTACTGCTCGGAGGCGCAAGCTCTTAAAGCGG 4141
QY 1381 GlnG1YThrProProProProProSerArgApsLeuThrGlnAlaTYrLYsThrGln 1400
DB 4142 GAGGACAGCTCGCCCGCCACGCGCTCACGGAGCTGACGAGGCTTCAAGACCCAG 4201
QY 1401 AlaLeuG1YProLeuLYsLeuLYsProAlaH1eG1YLeuValAlaThrValLYsG1Y 1420
DB 4202 GCCCTGGGCCCTTGAACTGAGAGCCGCCCATGAGAGGCTGTGTGCGCAGGTGAGAG 4261
QY 1421 AlaG1YArgSer11eH1eGln11eProArgGlnG1YLeuArgH1eThrProG1YLeuPro 1440
DB 4262 GCGGGCGGCTCCATCATGATCCCGGAGAGACTGTGGGACACGCGCGAGCTGCC 4321
QY 1441 LeuAlaProArgProLeuYsG1YSer11eThrGlnG1YThrProLeuLYsTYrAps 1460
DB 4322 CTGGCCCGGGCGGCTCAAGAGGGCTCCATCAGGAGGACCCCGCTCAAGTACGAC 4381
QY 1461 ThrG1YAlaSerThrThrG1YSerLYsLYsH1eApsValArgSerLeu11eG1YSerPro 1480
DB 4382 ACCGGGCGGTCCACCACTGGCTCAAAAAGCAGAGTAGTACCTCTCATCGGCAAGCCC 4441
QY 1481 G1YArgThrPheProProValH1eProLeuApsValMetAlaApsAlaArgAlaLeuG1Y 1500
DB 4442 GGCCTGAGCTTCCACCCGTGACCCGCTGATGTATGCGCAGCCCGGCACTGGAA 4501
QY 1501 ArgAlaCYeTYrGlnG1YSerLeuYsSerArgProG1YThralaSerSerG1YG1Y 1520
DB 4502 CGTGTGCTACAGGAGAGCTGAGAGCCGCGCAGGAGCGCAGAGCTGGGGGCG 4561
QY 1521 Ser11eAlaArgG1YAlaProVal11eValProG1YLeuG1YLYsProArgGlnSerPro 1540
DB 4562 TCCATTGGCGCGGCGCCCGGTGATGTGTGCTGAGACTGGTAAAGCCGCGCAGAGCCCC 4621
QY 1541 LeuThrTYrGlnApsH1eG1YAlaProPheAlaG1YH1eApsProArgG1YSerProVal 1560
DB 4622 CTGACCTTAGAGGACCAAGGAGACCTTGTGCGGCACTCCCAAGAGGTTGCGCGTG 4681
QY 1561 ThrMetArgG1YProThrProArgLeuGlnG1YSerLeuSerSerSerLYsAlaSer 1580
DB 4682 ACCATGCGGAGGCCACGCGCGGCTGACAGAGGGAGGCTTTCGTCAGCAAGGATCC 4741
QY 1581 GlnApsArgLYsLeuThrSerThrProArgGln11eAlaLYsSerProH1eSerThrVal 1600
DB 4742 CAGAGACCAAGACTAGCTGAGAGCTGTGATGATGSCAAAGTCCCGCAGACAGCAGTG 4801
QY 1601 ProG1YH1eH1eProH1aPro11eSerProTYrGlnH1eLeuLeuArgG1YAlaSerG1Y 1620
DB 4802 CCCGAGCACACCAACCCCATCTGCGCTATGAGACCTGTCTGGGGCGTGAAGTGC 4861

1621 ValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGly 1640
4862 GTGGACCTGTATTCGACGCACATCCCCCTGGCTTCGACCCACCTCCATACCCCGGCG 4921
1641 IleProLeuAspAlaAlaIleAlaIleTyrTyrLeuProArgHisIleAlaIleProAspProThr 1660
4922 ATCCCTGTGAGCGAGCCGCTGCTACTACTGCTGCCGACACCTGGCCCCCAACCCCAAC 4981
1661 TyrProHisIleLeuTyrProProTyrLeuIleArgGlyTyrTyrProAspThrAlaIleLeuGlu 1680
4982 TACCCGCACTGTACCCACCTCACTCATCCGGGTACCCGACACGGCGGGGTGGAG 5041
1681 AsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnIleMetHisIleAsnThrAla 1700
5042 AACCCGGACACCATCATCATCACTACATCACTCGCAGACAGATGACACCAACACGGCC 5101
1701 ThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeu 1720
5102 ACCGCACTGGCCAGCGAGCTGATATGTAGAGGGGCTCTCGCCCCGAGATCTCTGCTG 5161
1721 AlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnAlaProHisIleAla 1740
5162 GCACTCACTAGCTGGGGGTCCCGAGGCAATCATGACCTGTCCCAAGTGCACACCTG 5221
1741 ProValIleuValProProThrProGlyIleThrProAlaThrAlaMetAspArgLeuAlaTyr 1760
5222 CCTGTGCTGTGCCCCCGGACACCAAGCACCCGACCCGATGAGACCGCTGTGCTG 5281
1761 LeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGly 1780
5282 CTCGCCACCGCCGCCAGCCCTTCAGACACCGCCAGACAGCTCCCTCCACGAG 5341
1781 GlyProThrHisIleuThrTyrSerProThrThrThrSerSerSerGluArgGluArgAspArg 1800
5342 GGTCCAAACACTTGCACAAACCAACACACAGTCTCGTCCGAGGGAGGAGGAGCCGG 5401
1801 AspArgGluArgAspArgAspArgGluArgGluArgSerIleLeuThrSerThrThrThr 1820
5402 GATCCAGACCGGAGCCGGGATCGGAGCGGGAAATCCATCTCCATCCATCCACACAGAG 5461
1821 ValGluHisAlaProIleTyrParProGlyIleThrGlnIleSerSerGlySerSerGlySer 1840
5462 GTGGAGCAGCAGCAGCCATCTGAGACCTGTGACAGACAGACAGCGCGCAGCGCAGAC 5521
1841 SerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHis 1860
5522 ACGCGCGGGGTGGGGCAGCAGCAGCCGCGCTCCCATCTCCATGCCCCACAGCAC 5581
1861 SerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsn 1880
5582 TCGCCCATCTCCCTCGGACCCAGATGCCCTTCACAGACAGACCCAGTGTGCTTCACAA 5641
1881 ThrGlyMetLeuGlyIleIleThrAlaValGluProSerIleProThrValLeuArgSer 1900
5642 ACAGGCAATGAAGGATCATCAACGCTGTGAGCCAGACAGCCACGGTCCCTGAGAGTCC 5701
1901 ThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisGlyAspPro 1920
5702 ACCCTCACTCTCCACCCGTTGGCCAGCTGCACATCCACCTGCACCCACCTCCCA 5761
1921 LeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValIleLeuProGly 1940
5762 CTGGGGGGGACCTCCATGGGGGTCTACCCCTCATGGAGGCCGTCTTGCTGCCCAAG 5821
1941 GluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLeu 1960
5822 GAGGCCCCCGGGGTCCCGCGCAGAGCGGCCCGGAGCAGACACCGGCATGCTTCTC 5881
1961 AlaIleProProAlaArgSerGlyLeuGluProAlaSerSerProSerIleGlySerGlu 1980
5882 GCCAACCCCCCAACCCGCTCCGGGGCTGGAGCCCGCTCTCCCCACAGCAAGGGCTGGAG 5941

1981 ProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaIle 2000
5942 CCCCCGCTGTAGTCTCTGCTGTCTGTGACACGACCATGCGCCGACCCCTCCGAG 6001
2001 AsnLeuAlaProHisIleAlaSerProAspProProAlaProProAlaSerAlaSerAsp 2020
6002 AACCTTGACCTTCACACGACGACCGGAGCCCGCGGACCACTGGCTCGGCTCGGAC 6061
2021 ProHisArgGluIleThrGlnSerIleProPheSerIleGlnIleuGluLeuArgSer 2040
6062 CCGCACCGGGAAGAACTCAAGTAACCCCTTTCATCCAGGAACCTGGAACCTCGCTCT 6121
2041 LeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSer 2060
6122 CTGGGTATCAAGGACGAGCTACAGCCCGGAAAGGGGTGGAGCCCGTCAGCCCTGTGAC 6181
2061 SerProSerLeuThrHisAspArgGlyLeuProGlyLeuProGlyHisIleuGluIleuAspArgSer 2080
6182 TCACCCAGTGTGACCCACGACAAAGGGGCTCCCAAGCACCTGGAAAGAGCTCCAGAAAGAC 6241
2081 HisLeuGluGlyGluLeuArgProArgGlnProGlyProValIleLeuGlyGlyAla 2100
6242 CACCTGGAGGGGAGCTGGGCGCAAGACGACGAGCCCGGTGAGCTTGGCGGAGGCG 6301
2101 AlaHisIleuProHisIleuArgProLeuProGluSerGlnProSerSerSerProLeuLeu 2120
6302 GCCCACTTCCACACTGGCGCGCTGCTGTGAGACCAAGCCCTGTCCAGCCCGCTGCTC 6361
2121 GlnThrAlaProGlyValIleGlyHisGlnArgValIleThrLeuAlaGlnHisIleSer 2140
6362 CAGACCGGCCCGAGGGGTCAAGGTCAACAGCGGGTGTCACTTGCCACGACATCAAT 6421
2141 GluValIleThrGlnAspTyrThrArgHisIleProGlnIleuSerAlaProLeuPro 2160
6422 GAGGTATACACAGAGACTACACCGGACCAACACAGACACTGACGCGACCCCTGCC 6481
2161 AlaProLeuTyrSerPheProGlyAlaSerCysProValIleuAspLeuArgArgProPro 2180
6482 GCCCCCTCTACTCTCTCCCTGGGGCAGCTCCCGTCTGTGAACCTCGCGGCCACCC 6541
2181 SerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerProHisSer 2200
6542 AGTGACCTTACTCTCCGCGCCGAGACATGGTGGCCCGGTGCTCCGCCACAG 6601
2201 GluGlyGlyLeuArgSerProGluProAsnIleThrSerValIleuGlyGlyGluAsp 2220
6602 GAAAGGGGGAAGGCTCTCAGAGCAAGACGTCGCTGTTGGGTGTGGTGAAGAC 6661
2221 GlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAla 2240
6662 GGATATTGAACCTGTGTCCCAACCGAGGGGCAATGACGAGCCAGGCACTCCCGAGTGTCT 6721
2241 ValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerIle 2260
6722 GTGTACCGCTGTGACCGGATGGGAAACAGACGAGCCACAGACAGATGGGCTCCAA 6781
2261 SerProGlyAsnThrSerGlnProProAlaPhePheSerIleLeuThrGluSerAsnSer 2280
6782 TCTCAGGCAACACGACGACGCGCAGCTTCTTCAGCAACCTGACGAGACACTCC 6841
2281 AlaMetValIleSerIleArgGlnGluIleAsnIleGlyLeuAsnThrHisAsnAlaGln 2300
6842 GCCATGTCAAGTCCAAAGAGAGATCAACAGAACTTAACCCCAACACCGGAAT 6901
2301 GluProGluTyrAsnIleSerGlnProGlyIleThrGluIlePheAsnMetProAlaIleThr 2320
6902 GAGCTGGAATACATATACGACGCTGGAGAGAGATCTTCAATATGCCCGCAATACC 6961
2321 GlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisIleAsnThrAsnMet 2340
6962 GGAACAGGCTTATATGACTATATGAAGCCAGCGGTGCAAGAAATATCCAGACCAACATG 7021
2341 GlyLeuGluAlaIleIleIleArgIleAlaLeuMetGlyIleTyrAspGlnIleProGluGluSer 2360

|||||
Db 7022 GGGCTGAGGCGCATATAGAAAGCACTCATGGGTAAATATGACCAAGTGGGAAGAGTCC 7081
Qy 2261 ProProLeuSerAlaAaPheAsnProLeuPheAsnAlaSerAlaSerLeuProAlaAla 2380
Db 7082 CCGCGGCTCAGCGCAATGCTTTAACTCTGATGACGATGCGAGCTTCGCGTCT 7141
Qy 2381 MetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGly 2400
Db 7142 ATGCCCATTAACCGTGTGCGAGCGAGTGAACCACTCCTGCGCAAGTGGCGGC 7201
Qy 2401 GlyValAlaValSerGlyArgProSerSerArglyAlaValSerProAlaProGly 2420
Db 7202 GGGAGGCGCAAGGCTCTGTGCGAGACCAAGCAAGCAAGTCCCGCGCGCGC 7261
Qy 2421 LeuAlaSerGlyAspArgProProSerValSerSerValHisSerGlyGlyAspCysAsn 2440
Db 7262 CTGGCATCTGGGAGCGCGCACCTCTGTCTCTCAGTGCATCTGGAGGAGACTGCAAC 7321
Qy 2441 ArgArgThrProLeuThrAsnArgValTropGluAspArgProSerSerAlaGlySerThr 2460
Db 7322 CGCGGAGCGCGCTCACCAACCGCGTGTGGAGAGACAGCCCTGCTCCGCAAGTTCCACG 7381
Qy 2461 ProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProPro 2480
Db 7382 CCATTCCTCCATCAACCCCTGATCATGCGGCTGCAGCGGGGTTCATGTGCTTCCCAACC 7441
Qy 2481 ProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisAlaAlaTyrAspGlu 2500
Db 7442 CCACCGGCGCTCCCCGGGCGAGCGGGCCCTCTGCTGCGCCCAACCGCTGGGAGCAG 7501
Qy 2501 GluProLysProLeuLeuCysSerGlnTyrGluThrLeuSerAspSerGlu 2517
Db 7502 GAGCCCAAGCCTACTGCTCTGCTCCGATACGAGCACTCTCCGACAGCAG 7552

RESULT 3
US-11-052-344-1
; Sequence 1, Application US/11052344
; Publication No. US20050191674A1
; GENERAL INFORMATION:
; APPLICANT: Cognate, Lionel
; TITLE OF INVENTION: Method Of Prognosis Of Metastasis By Detection Of FRA12E Fragile
; TITLE OF INVENTION: Within The SMRT Gene/Locus At Chromosome 12q24
; FILE REFERENCE: 03551.0181
; CURRENT APPLICATION NUMBER: US/11/052,344
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/542,538
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 1
; SEQ ID NO 1
; LENGTH: 8561
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: SMRT gene which includes the open reading frame for SMRT protein.
; OTHER INFORMATION: The open reading frame starts at nucleotide 2 and ends at nucleot
US-11-052-344-1

Alignment Scores:
Pred. No.: 0 Length: 8561
Score: 13215.00 Matches: 2517
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 26 Gaps: 0

US-09-522-753-5 (1-2517) x US-11-052-344-1 (1-8561)

Qy 1 MetSerGlySerThrGlnLeuValAlaGlnThrTyrArgAlaThrProArgTyrPro 20
Db 2 ATGTGGGCTCCACACAGCTTGTGGCACAGCGTGGGCGCACGTGAGCCCGCTACCG 61
Qy 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40

|||||
Db 62 CCCACAGCCTTCTCTACCCAGTGCAGATGCCCGGAGCACAGCAAGTGGGCTCTCTG 121
Qy 41 GlnTyrGlnHisSerArgAspTyrAlaSerHisLeuSerProGlySerTlleGln 60
Db 122 GAGTACAGCACCACTCCCGGACTATGCTCCCACTGTGCGGGGTTCATATCCAG 181
Qy 61 ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
Db 182 CCCAGGCGCGAGGCCCTCTCTGTCTGTAGTTCAGCCCGGGAATGAACGGTCCAG 241
Qy 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlySerGluMet 100
Db 242 GAGTCCACCTCGCGGCGAGAGTCCCATCTACCTGCCCGAGCTGGGGAATCAGAGATG 301
Qy 101 GluPheIleGlySerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
Db 302 GAGTTCAATTAAAGCAAGCCCTCGGCTAGAGTGTGCTGCTGACCCCTGTGCGACCG 361
Qy 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
Db 362 TCACCTCTGCTGCGCACGGGCGAGCTGCGGATCTAAGACTCACCAAGAACCTGATGC 421
Qy 141 LeuThrGlyLysLeuGluProValSerProProSerProHisThrAspProGluLeu 160
Db 422 CTGACGGGCAAGCTGGAACCGGTGTCTCCCGCACGCCCGGACACTGACCTGAGCTG 481
Qy 161 GluLeuValProProAlaGlyLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAsp 180
Db 482 GAGCTGTGTGCGCGCACCGGCTGTCCAGAGAGCTGTATCCAGAACATGAGACCGGTGGAC 541
Qy 181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnLeu 200
Db 542 CGAGAGTCAACCATGTGATGAGACAGACATCTTAAGCTGAAGAAAGACACACAGCTG 601
Qy 201 GluGluGluAlaAlaLysProProGluProGluLysProLysSerProProIleGlu 220
Db 602 GAGGAGAGGCTGCGCAAGCCCGGAGCTGAGAGAGCCCGTGTACACCGCGCCATCGAG 661
Qy 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysGluAla 240
Db 662 TCAGAGCACCGCAGCCCTGTGAGATCATCTACGACAGAACCGGAGAGGCTGAGAGCT 721
Qy 241 AlaHisArgLleLeuGluGlyLysGlyProGluValGluLeuProLeuTyrAsnGlnPro 260
Db 722 GCACTGTGATTTCTGAAAGCTTGGGCGCCCAAGTGAAGTCCGCTGTACACCAAGCCC 781
Qy 261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLeu 280
Db 782 TCCGACACCCGGCAGTATCATGAAACATCAAAATAAACAGCGGAGTGGGAAAGACTA 841
Qy 281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGln 300
Db 842 ATCTTGACTTCAAGAGAGGAATCAAGCTCGGAAACAAATGGAAGCGAAGTTCTGCGAG 901
Qy 301 ArgTyrAspGlnLeuMetGluAlaLeuGluLysValGluArgIleGluAsnAspPro 320
Db 902 CGCTATGACCAAGCTCATGAGGCTTGGAAABAAAGGTGAGGACATCGAAACACCGC 961
Qy 321 ArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340
Db 962 CGCGGCGGGCGCAAGGAGAGCAAGTGGCGGAGTACTACGAAAGAGCTTCCGTGAGATC 1021
Qy 341 ArgLysGlnArgGluLeuGlnGlnLysMetGlnSerArgValGlyGlnArgLysGly 360
Db 1022 CGCAACAGCGCAGCTGCAGAGCCATGCAAGACAGCAGGCTGGGCCACGCGGCAAGTGG 1081
Qy 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluTyrIleIleAspGlyLeuSer 380
Db 1082 CTGTTCATGTGCGCGCGCCCGGACGACGACAGAGTGTCAAGATCATATGCTCTCA 1141
Qy 381 GluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProPheMetLeuTyr 400

Dh 1142 GAGCAGAGAACCTGGAGAGCAGATGCCGACGCTGGATCCCGCCATCGTGTAC 1201
Qy AASPAlaAspGlnGlnArgIleIysPheIleAsnMetAsnGlyLeuMetAlaAspPromet 420
Dh 401 |||||
Db 1202 GAGCGTACAGAGGAGGATCAAGTTTCATCAACATAGACGGCTTATGGCCGACCCATG 1261
Qy 421 LysValaTyIysAspArgIleValMetAsnMetTrpSerGluGlnIlyysGluThrPhe 440
Dh 1262 AAGGTGTACAAAGACCGCCAGGTCAATGACATGTGGATGAGCAGAGGAAGAGACTTTC 1321
Qy 441 ArgGluIlyPheMetGlnIleSpProIyAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
Dh 1322 CGGGAAAGTTCAATGACGATCCCAAGAACTTTGGCTGATCCCATTCATTCCCGAGAG 1381
Qy 461 LysThrValAlaGluCysValLeuTyIysTyIysThrIlyPheAsnGluAsnTyIys 480
Dh 1382 AAGACGTGGCTGAGTGCCTCTCTTACTTACTGACTAGAGAAATGAGAACTTAAAG 1441
Qy 481 SerLeuValArgArgSerTyIysArgArgIlyysSerGlnGlnGlnGlnGln 500
Dh 1442 AGCTGTGAGAGAGGATATCGGCCCGGCAAGGACGACGACGACGACGACGACGACG 1501
Qy 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
Dh 1502 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1561
Qy 521 AAspGluIlySerGluIlySerGluIlySerGluIlySerGluIlySerGluIlySer 540
Dh 1562 GATGAG 1621
Qy 541 AAspIlySerGluIlySerGluIlySerGluIlySerGluIlySerGluIlySer 560
Dh 1622 GACAGAGAGAGACTCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1681
Qy 561 LysGluAlaValAlaSerIlySerGluIlySerGluIlySerGluIlySerGluIly 580
Dh 1682 AAGGAGAGCTGTGCTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1741
Qy 581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluIlyAlaIleThrProGln 600
Dh 1742 CGCATCAACCGCTCAATGGCTTAATGAGGCCAAGCAGAGAGAGAGAGAGAGAG 1801
Qy 601 SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluMet 620
Dh 1802 AGGCGCAGCTGCGCTCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1861
Qy 621 GluThrAlaIlySerGluLeuLeuGluIlySerGluIlySerGluIlySerGluIly 640
Dh 1862 GAAACAGCCCAAGAAAGGTCTCTGGAACAGCGCGCAACTGTGCGGCATCGCCGATG 1921
Qy 641 ValGlySerIlyThrValSerGlnCysIlyAsnPheTyIysPheAsnTyIlySerGln 660
Dh 1922 GTGGGCTCCAGAGCTGTGTGCAGTGTAAAGACTTCTTCACTTCACTTCAAGAGAG 1981
Qy 661 AsnLeuAspGluIleLeuGlnGlnIlySerGluIlySerGluIlySerGluIlySer 680
Dh 1982 AACCTCGATGAGATCTTGCAGCAGCAGCAACAGCTGAATGAGAGAGAGAGAGAG 2041
Qy 681 ArgIlySerIlySerGluIlySerGluIlySerGluIlySerGluIlySerGluIly 700
Dh 2042 AGGAAAG 2101
Qy 701 GluAspGluGluMetGluIlySerGluIlySerGluIlySerGluIlySerGluIly 720
Dh 2102 GAGGATGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2161
Qy 721 AlaGluAlaLeuIleAlaSerGluIlySerGluIlySerGluIlySerGluIlySer 740
Dh 2162 GGTGAAGCTTACATGCTCTGGGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2221
Qy 741 ThrValAsnAsnSerSerAspThrGluSerIleProSerProIleSerGluAlaIly 760
Dh 2222 ACTGTCAACAAAGCTCAGACCGAGAGATCCCTCTCTCTCACTGAGGCGCGCAAG 2281

Qy 761 AspThrGluIlyIlyAsnGluIlyProIlySerProAlaIleThrLeuGluAlaAspGluIlyProPro 780
Dh 2282 GACACAGGCGAGATGGGCGCAAGCGCCAGAGCACTCTGGGCGCGAGAGAGAGAGAGAG 2341
Qy 781 GlyProProThrProProArgArgThrSerArgAlaProIleGluIlyProThrProAlaSer 800
Dh 2342 GGGCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 2401
Qy 801 GluAlaThrGluAlaProThrProProProAlaProProSerProSerAlaProProPro 820
Dh 2402 GAAGCCACCGAG 2461
Qy 821 ValIlyProIlySerGluIlySerGluIlySerGluIlySerGluIlySerGluIly 840
Dh 2462 GTGTCTCCCAAG 2521
Qy 841 GluGluGluIlySerProProAlaAlaGluGluIlySerGluIlySerGluIlySer 860
Dh 2522 GAGGAGCAG 2581
Qy 861 ProValIlySerGluIlySerGluIlySerGluIlySerGluIlySerGluIlySer 880
Dh 2582 CCGGTCAAG 2641
Qy 881 GluAlaAlaGluAlaThrAlaGluIlyAlaLeuIlySerGluIlySerGluIlySer 900
Dh 2642 GAGGCGCTGAGGCGCAGGCGCAGGCGGCGCTCAAGAGCAGAGAGAGAGAGAGAGAG 2701
Qy 901 GlyAlaAlaThrThrAlaIlySerSerGluIlyAlaProGluIlySerSerSerAlaThr 920
Dh 2702 GGCAGGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2761
Qy 921 CysSerAlaAspGluIlyAlaAspGluIlySerGluIlySerGluIlySerGluIly 940
Dh 2762 TGCAGTGCAG 2821
Qy 941 ArgProSerLeuIlyThrProThrGluIlyAspProArgAlaAsnAlaSerProGluIlyPro 960
Dh 2822 AGGCGCAGCTCTCAACCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2881
Qy 961 LeuAspLeuIlySerGluIlySerGluIlySerGluIlySerGluIlySerGluIly 980
Dh 2882 CTGGACCTTAAG 2941
Qy 981 ValIleGluProProArgGluAspAlaAlaProThrIlyProAlaProProAlaProPro 1000
Dh 2942 GTTCATGAG 3001
Qy 1001 ProProGluIlyAsnLeuGluIlyProGluIlySerAspAlaProGlnGluIlySerSerProArg 1020
Dh 3002 CCAACCGCAAAACTGTCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3061
Qy 1021 GlyIlySerSerArgSerProAlaProProAlaAlaIlySerGluIlySerGluIly 1040
Dh 3062 GGCMAAG 3121
Qy 1041 LysLeuProGluIlyAspProProCysTrpThrSerGluIlyLeuProPheProValProProArg 1060
Dh 3122 AAGCTGCTGTGGAG 3181
Qy 1061 GluValIleIlySerAlaSerProIleAlaProAspProSerAlaPheSerTyIlyAlaProPro 1080
Dh 3182 GAGGTATCAAGGCTCTCCCGCATGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3241
Qy 1081 GlyIleProLeuProLeuGluIlyLeuIleAspThrAlaArgProValLeuProArgProPro 1100
Dh 3242 GGTCAACCACTGCGCTGTGGCTCTCATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3301
Qy 1101 ThrIleSerAsnProProProLeuIleSerSerAlaIlySerProSerValLeuGluArg 1120
Dh 3302 ACCATCTCAACCGGCTCTCTCATCTCTGTGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 3361

QY 1121 GlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGlnHis 1140
DB 3362 CAATATAGTGCATCTCCCAAGAAATGTCGGTCCAGCTCCACGTCCTCCACTGAGACAT 3421
QY 1141 AlAluValAProValGlyProValThrmGlyLeuProLeuProMetAAsProLys 1160
DB 3422 GCCAAGGCCCCGGGGGCGCTGTCACATGGGGCTGCCCTGCTCAATGAGACCCCAAAAG 3481
QY 1161 LeuAlaProPheSerGlyValLysGlnGlnLeuSerProArgGlyGlnAlaGlyPro 1180
DB 3482 CTGGACCCCTTCACGGAGTGAAGCAAGACAGCTGTCCTCCAGGGGCGAGCTGGGCCA 3541
QY 1181 ProGlnSerLeuGlyValProThraIleGlnLysAlaSerValLeuArgGlyThraIleu 1200
DB 3542 CCGGAGAGCTGGGGGGTGGCCACAGCCCAAGAGGCGTCGCTGAGAGGAGACAGCTCTG 3601
QY 1201 GlySerValProGlyGlySerIleThrLysGlyIleProSerThraArgValProSerAAs 1220
DB 3602 GGCCTAGTTCGGGGCGGAGCATCACAAAGGCAATCCCGACACACAGGAGCTCCGAGC 3661
QY 1221 SerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAAsValIleuTyrLys 1240
DB 3662 AGCGCCATCAATACCGGGCTCATCACACGCGACGCGAGCTGACGTCCTGACAG 3721
QY 1241 GlyThrIleThraArgIleIleGlyLysAAsProSerArgLeuAAsArgGlyArgGln 1260
DB 3722 GGCACCATCACAGATCATCGGAGAGACAGCCGAGTGGCTTGACCGGGGCGGAG 3781
QY 1261 ArgSerLeuProLysGlyHisValIleTyrGlnGlyLysLysGlyHisValLeuSerTyr 1280
DB 3782 GACAGCTGCGCCAGAGGCGCACGTCATCTACAAAGGCAAGAGGCGCCGCTTGTCTAT 3841
QY 1281 GlnGlyGlyMetSerValThrGlnCysSerLysGlnAAsArgLysSerSerGlyPro 1300
DB 3842 GAGGAGTGCATGCTGTGACCAAGTGTCCAAAGAGACGCGCAAGACACTCAGAGCC 3901
QY 1301 ProHisGlyThraIleAAsProLysArgThrTyrAAsMetMetGlnGlyArgValGlyArg 1320
DB 3902 CCCCATGAGACGGCGCGCCCAAGCGCATGATGACATGATGAGGGCGCGTGGGAGA 3961
QY 1321 AlaIleSerSerAlaSerIleGlyGlyLeuMetGlyArgAlaIleProProGlnLysHis 1340
DB 3962 GCCATCTCCACACGACCATCGAAGCTCATGGGCGCTCCATCCCGCGAGCGCAC 4021
QY 1341 SerProHisIleuLysGlnGlnHisIleArgGlySerIleThrGlnGlyIlePro 1360
DB 4022 AGCCCCACCACTTCAAGACAGCACACATCCGCGAGCTCATCACAAAGGATCCT 4081
QY 1361 ArgSerTyrValGlnLysGlnAAsProTyrLeuArgArgGlnAlaLysLeuLeuLysArg 1380
DB 4082 CGGTCTTACGTGAGGACAGAGGACTACTGTCGGAGGCGCAAGCTCTTAAAGCGG 4141
QY 1381 GlnGlyThrProProProProProProProProProProProProProProProPro 1400
DB 4142 GAGGAGACGCTCCGCCCCCAGCGCTCCAGGAGCTGACCGAGGCTTCAAGAGCGCAG 4201
QY 1401 AlaLeuGlyProLeuLysLeuLysProAlaHisGlnGlyLeuValAlaThrValLysGln 1420
DB 4202 GCCCTGGCCCCCTTGAAGCTGAAGCGGCGCCATGAGGGCGCTGGGGCGAGGAGAG 4261
QY 1421 AlaGlyArgSerIleHisGlnIleProArgGlnGlnLysArgHisThrProGlnLeuPro 1440
DB 4262 GCGGGCGGCTCATTCATGAGATCCGCGGAGAGCTGGGACACGCGCGAGCTGGCC 4321
QY 1441 LeuAlaProArgProLeuLysGlnGlySerIleThrGlnGlyThrProLeuLysTyrAAs 1460
DB 4322 CTGGGCCCGCGCGCTCAAGAGGAGCTCATCACAGAGGCGACCCGCTCAAGTACGAC 4381
QY 1461 ThrGlyAlaSerThrThrGlySerLysLysHisAAsValArgSerLeuIleGlySerPro 1480
DB 4382 ACCGGGCGCTCCACACTGGCTCCAAAGACAGAGCTTACGCTCCCTCATGGGAGGCC 4441
QY 1481 GlyArgThrPheProProValHisProLeuAAsValMetAlaAAsAlaArgAlaLeuGln 1500

DB 4442 GCGCGAGCTTCCACCGCTGACACCGCTGATGTATGACCGACGCGCGGACCTGAA 4501
QY 1501 ArgAlaCysTyrGlnGlySerLeuLysSerArgProGlyThraIleSerSerGlyGly 1520
DB 4502 CGTGCCTGTACAGAGGAGACTGTGAAGAGCGGCGCAAGGACCGCACACTCGGGGGG 4561
QY 1521 SerIleAlaArgGlyAlaProValIleValProGlnLeuGlyLysProArgGlnSerPro 1540
DB 4562 TCCATTCGGCGGGCGCGCGCTGATTTGCTGAGCTGGGTAAAGCGCGGAGAGCTCC 4621
QY 1541 LeuThrTyrGlnAAsHisGlyAlaProPheAlaGlyHisLeuProArgGlySerProVal 1560
DB 4622 CTAGACTATGAGACCAAGGGGACCTTTGGCGGACACTCCCAAGAGTTCGCGCTG 4681
QY 1561 ThrmAAsGlnProThrProArgLeuGlnGlySerLeuSerSerSerLysValAAs 1580
DB 4682 ACCATGGGAGCGCCACGCGCGCTGCAAGAGGCGACCTTTCGTCACCAAGGCAATCC 4741
QY 1581 GlnAAsArgLysLeuThrSerThrProArgGlnIleAlaLysSerProHisSerThraI 1600
DB 4742 CAGGACGAAAGCTGACGTGACGCTCGTGAATGCGCAAGTCCCGCACAGACCTG 4801
QY 1601 ProGlnHisIleProHisProIleSerProTyrGlnHisLeuLeuArgGlyLysGly 1620
DB 4802 CCGGAGACCAACCAACCCCATCTGCGCTTATGAGACACTGCTGGGGCGTGAAGTGC 4861
QY 1621 ValAAsLeuTyrArgSerHisIleProLeuAlaPheAAsProThrSerIleProArgGly 1640
DB 4862 GTGACCTGTATGAGCGCCATCCCTGGCTTGACCCCACTCCATACACCCCGGCG 4921
QY 1641 IleProLeuAAsAlaIleAlaIleTyrTyrLeuProArgHisIleuAlaProAAsProThr 1660
DB 4922 ATCCCTTGGACCGACCGCTGCTTACTACTGCGCGACACTTGCGGCCCAACCCAC 4981
QY 1661 TyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAAsThraIleuGln 1680
DB 4982 TACCCGACCTGATACCAACCTTACTCATCTCGGGGCTACCCGACAGCGGGCGCTGAG 5041
QY 1681 AsnArgGlnThrIleIleAsnAAsProTyrIleThrSerGlnGlnMetHisIleAAsThraI 1700
DB 5042 AACCGGACAGCACTCATGATGATCATCTCCAGAGATGACCAACCAACCGGCC 5101
QY 1701 ThraIleMetAlaGlnAAsArgAlaAAsMetLeuArgGlyLeuSerProArgGlnSerSerLeu 1720
DB 5102 ACCGCAATGGCCAGAGAGCTGATGCTGAGGGGCTCTCGCGCGCGAGTCTCGCTG 5161
QY 1721 AlaLeuAAsTyrAlaIleGlyProArgGlyIleIleAAsLeuSerGlnValProHisLeu 1740
DB 5162 GCACTCAACTACGCTCGGGTCCCGAGGACATGACCTGTCCCAAGTGCACACCTG 5221
QY 1741 ProValLeuValProProThrProGlyThrProAlaThraIleMetAAsArgLeuAlaTyr 1760
DB 5222 CTGTGCTGTGGCCCCCGACACCAAGGACCCCAAGCGCTTGAAGCGCTTGGCTTAC 5281
QY 1761 LeuProThraIleProGlnProPheSerSerArgHisSerSerSerProLeuSerProGly 1780
DB 5282 CTCCCAACGCGCCAGGCTTTCAGACCGGCGCACAGAGCTCCCACTTCCCAAGGA 5341
QY 1781 GlyProThrHisIleuThrLysProThrThrThrThrSerSerSerGlnArgGlnAAsArg 1800
DB 5342 GGTCAACACACTTGAACAAACCAACACACAGTCTCGTCCGAGCGGAGCGAGACCGG 5401
QY 1801 AAsArgGlnAAsArgAAsArgAAsArgGlnGlnLysSerIleLeuThrSerThrThrThr 1820
DB 5402 GATCGAAGACCGGAGCGGAGTGGGAGCGGAAATCATCTTCACTCCACACAGACG 5461
QY 1821 ValGlnHisAlaProIleThraArgProGlyThrGlnGlnSerSerGlySerSerGlySer 1840
DB 5462 GTGGAGACGACCACTTGGAGACTGTGATCAGACAGACAGCGGCGAGCGGCGAGC 5521
QY 1841 SerGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHis 1860

Db 5522 AGCGGCGGGGTGGGGGACAGACCGCCCGCTCCACTCCCATCCACGACGAC 5581
Qy 1861 SerProIleSerProArgThrGlnAspAlaLeuGlnArgProSerValLeuHisAsn 1880
Db 5582 TGGCCATCTCCCTCGAGCCAGGATGCCCTCCACAGAGACCCAGTGTCTTCACAC 5641
Qy 1881 ThrGlyMetIleGlyIleIleThrAlaValGluProSerIleAspThrValLeuArgSer 1900
Db 5642 ACAGGATAGAGGGATCATCACCGCTGTGAGACCCAGACGACCCAGGCTCGAGAGTCC 5701
Qy 1901 ThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCysPro 1920
Db 5702 ACCCTCACCTCTCCACCGCTTGCCACAGCTGCACATCTCCACCTGCCACCACTGCCCA 5761
Qy 1921 LeuGlyGlyThrLeuAspGlyValIleProThrLeuMetGluProValLeuLeuProLys 1940
Db 5762 CTGGGGGGGACCTCGATGGGGGTCTACCTCATGGAGCCGCTTGTGCTGCCAAG 5821
Qy 1941 GluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLeu 1960
Db 5822 GAGGCCCCCGGGGTGCGCGGCGAGCGGCGCCGACAGACACCGGCGCATGCTTCTC 5881
Qy 1961 AlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerIleGlySerGlu 1980
Db 5882 GCCAAGCCCCCAGCCGCTCGGGCTGAGCCGCGCTCTCTCCCGACAGAGGGCTCGAG 5941
Qy 1981 ProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLys 2000
Db 5942 CCCCCGCGCTTGTGCTCTCTGTCTGTGGCAGCAGCCATGCGCGGACCTCGGGAAG 6001
Qy 2001 AsnLeuAlaProHisAlaSerProAspProProAlaProProAlaSerAlaSerAsp 2020
Db 6002 AACCTGCACCTCACACGACGACCGGAGCCGCGCGGCGCACCTGCTCGGCTCGAGC 6061
Qy 2021 ProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSer 2040
Db 6062 CCGCACCGGGAAAGACTCAAGATAACCTTTTCATCCAGGAACTGGAACTCCCTTCT 6121
Qy 2041 LeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSer 2060
Db 6122 CTGGGTTACCGGCGACGACTACAGCCCCGAGAGGGGTGAGCCCTGACCCCTGTGAGC 6181
Qy 2061 SerProSerIleThrHisAspLysGlyLeuProLysHisLeuGluLysLeuAspLysSer 2080
Db 6182 TCAACCAAGCTGACCCACGACGAGGGGCTCCCAAGACCTGGAAGAGCTCGACAGAGC 6241
Qy 2081 HisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGlyAla 2100
Db 6242 CACCTGGAGGGGAGCTGGCGCCAGACGACGCGCCGCTGAAGCTTGGCGGGAGGCC 6301
Qy 2101 AlaHisLeuProHisLeuArgProLeuArgProGluSerGlnProSerSerSerProLeuLeu 2120
Db 6302 GCCCACCCTCCACACTTGGGCGCGCTGCTGAGAGCCAGCCCTGCTCCAGCCCGCTGCTC 6361
Qy 2121 GlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSer 2140
Db 6362 CAGACCGGCCCAAGGGGTCAAGGTCAACAGCGGTGTGTCACCTGGCGCCAGCATCAAGT 6421
Qy 2141 GluValIleThrGlnAspTyrThrArgHisAspProGlnGluLeuSerAlaProLeuPro 2160
Db 6422 GAGGTATATCACAGACTACACCCGCGACACCCACAGAGTCAAGCGACCCCTGCGCC 6481
Qy 2161 AlaProLeuTyrSerPheProGlyValAspCysProValLeuAspLeuArgArgProPro 2180
Db 6482 GCCCCCCCTCTACTCTTCTCTGGGGCCAGGTCGCCGTCTTGAAGCTTCCGCCGCCACCC 6541
Qy 2181 SerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerProHisSer 2200
Db 6542 AGTGACTCTACTCCCGCGCCCGGACCATGTGCCCCGCGCTGTGCTCCCCACAGC 6601
Qy 2201 GluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyAlaAsp 2220
Db 6602 GAAGGGGGCAAGGAGTCTCCAGAGCCAAACAGACGTGGTCTTGGGTGTGTGAGAC 6661

Qy 2221 GlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAla 2240
Db 6662 GGTAATTGAACCTGTCTCCCAACCGGAGGGCATGACGAGACCGAGGACATCCCGGAGTGCT 6721
Qy 2241 ValTyrProLeuLeuTyrArgAspGlyGlyGlnThrGluProSerArgMetGlySerLys 2260
Db 6722 GTGTACCCGCTGTGTACCGGGATGGGAAACGAGCGAGCCAGAGGATGGGCTTCAG 6781
Qy 2261 SerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSer 2280
Db 6782 TCTCCAGGCACACACGACGCCGCCAGCTTCTTCAGAGACTACCGAGGCACTCC 6841
Qy 2281 AlaMetValLysSerLysGlyGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsn 2300
Db 6842 GCCATGTCTAACTCCAGAGACGAGATCAACAGAGAGCTGAACCCACCAACCGGAAT 6901
Qy 2301 GluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThr 2320
Db 6902 GAGCTGTAATACAAATATCAGCCAGCCTGGAGAGAGATCTTCAATATGCTCCCATCAC 6961
Qy 2321 GlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMet 2340
Db 6962 GGAACAGGCTTATGACCTATGAAAGCCAGCGGTGCAAGAACATGCCAGCACCAACATG 7021
Qy 2341 GlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTyrGluGluSer 2360
Db 7022 GGGCTGAGAGCCATATATGAAAGCACATCATGGGTATATATGACCATGGGAGAGTCC 7081
Qy 2361 ProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAla 2380
Db 7082 CCGCGCTCAGGCGCAAGTCTTTTAACTCTGAAATGCCAGTGCCAGCTGCCGCTGCT 7141
Qy 2381 MetProIleThrAlaAlaAspGlyArgSerAspHisIleThrLeuThrSerProGlyGly 2400
Db 7142 ATGCCATTAACCGCTGTCTGACGAGAGTACACACACTCAGCTCCGCAAGTGGCGGC 7201
Qy 2401 GlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGly 2420
Db 7202 GGGAAAGCCCAAGTCTCTGGCAGACCCAGCACCGGAAAGCCCAAGTCCCGCGCGCGC 7261
Qy 2421 LeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluGlyLysAspAsn 2440
Db 7262 CTGGCATCTGGGAGACCGGACCTCTGTCTCTCAAGTCACTCGAGGAGACCTGCAAC 7321
Qy 2441 ArgArgThrProLeuThrAsnArgValTyrGluAspArgProSerSerAlaGlySerThr 2460
Db 7322 CCGCGGAGCGCGCTCACCAACCGCGTGGAGAGACAGGCCCTCGTCCGCAAGTTCCAG 7381
Qy 2461 ProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProPro 2480
Db 7382 CCATTCCTCTAACACCCCTGATCATGCGGCTGACAGGCGGGTGTCTATGCTTCCACCC 7441
Qy 2481 ProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaIleTrpAspGlu 2500
Db 7442 CCAACCGGCGCTCCCGCGGAGCGGGCCCTCGCTGCGGCCACCAACCTGGAGCGAG 7501
Qy 2501 GluProLysProLeuLeuCysSerGlnTyrGluThrLeuSerAspSerGlu 2517
Db 7502 GAGCCCAAGCACTGCTGTCTGTGCAAGTACGAGACACTCTCCGACAGCGAG 7552

RESULT 4
US-10-087-192-653
; Sequence 653, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01

PRIOR APPLICATION NUMBER: US 09/747,377
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/798,586
PRIOR FILING DATE: 2001-03-02
NUMBER OF SEQ ID NOS: 2059
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 653
LENGTH: 8667
TYPE: DNA
ORGANISM: Homo sapiens
US-10-087-192-653

Alignment Scores:
Pred. No.: 0 Length: 8667
Score: 13145.00 Matches: 2509
Percent Similarity: 99.68% Conservative: 1
Best Local Similarity: 99.64% Mismatches: 7
Query Match: 99.47% Indels: 2
DB: 14 Gaps: 0

US-09-522-753-5 (1-2517) x US-10-087-192-653 (1-8667)

QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpAlaGlnAlaThrGluProAlaGlyTyrPro 20
DB 116 ATGTGGGATCCACACAGCCCTGTGGCACAGCTGGAGGGCCACTGAGCCCCGCTACCCG 175
QY 21 ProHisSerLeuSerTyrProValGlnLeuAlaArgThrHisSerProValGlyLeuLeu 40
DB 176 CCCACAGCTTTCTTCCATCCAGTGCAGATGCCCGAGCGACACAGCGATCGGGCTCTG 235
QY 41 GluTyrGlnHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
DB 236 GAGTACAGACACCACTCCCGAGTATGCTCCACCTGTGCGGCTCCATCATCAG 295
QY 61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
DB 296 CCCACAGGGGAGGGCCCTCTGCTGTCTGAGTTCCAGCCCGGGAATGAACGGTCCAG 355
QY 81 GluLeuHisLeuArgProGlnSerHisSerTyrLeuProGlnLeuGluYlySerGlnMet 100
DB 356 GACCTCCACTGGGGCCAGAGTCCCACTCATCTGCCAGCTGGGGAAGTCAGAGATG 415
QY 101 GluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
DB 416 GACTTATTGAAGAAGCAAGCCCTCGGCTAGAGTGTGCTGACCCCTCGTGCAGCG 475
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
DB 476 TCACCCCTGTGGCCACGGGCGAGCCCTGCGGATCTGAMAGACTTCACCAAGGACCGTAGC 535
QY 141 LeuThrGluYlyLeuGluProValSerProProSerProHisSerProGlnLeu 160
DB 536 CTACAGGGAGAGCTGGAAACGGGTGTCTCCCCAGCCCCCGGACACTGACCTGAGCTG 595
QY 161 GluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAsp 180
DB 596 GAGCTGTGCGCCACAGGCTGTCCAAAGAGAGACTATCCAGACATGAGCCCGCTGCAGC 655
QY 181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnLeu 200
DB 656 CGAGAAATCCACCTGTAGAGCAGAGATCTTAAGCTGAAGAAGAACGACGACGCTG 715
QY 201 GluGluGluAlaAlaLysProProGluProGluLysProValSerProProIleGlu 220
DB 716 GAGGAGAGAGCTGCCAAGCGCCGAGCTGAGAAAGCCGCTGTACCGCCCATCTGAG 775
QY 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAla 240
DB 776 TCAGACACACGAGCCGTGGTGAAGATCATACGAGAGAAACCGAAGAGGTGAAGCT 895
QY 241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrTrpAsnGlnPro 260
DB 836 GCACATCGGATTCTGAAAGGCTGGGGCCCGAGGTGAGCTGCCGCTGTACCAACGAGCC 895

QY 261 SerAspThrArgGlnTyrHisGluAsnIleLysIleLeuAsnGlnAlaMetArgLysLeu 280
DB 896 TCCGACACCCGGCAGATTCATGAGACATCAAAATAAACAGGGAGATCCGGAAGACCTA 955
QY 281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGln 300
DB 956 ATCTTGACTTCAAGAGAGAAATCAACGCTCGGAAACAAATGGGAGCAGAAAGTTCTGCAG 1015
QY 301 ArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAspPro 320
DB 1016 CGCTATGACCACTCATGAGAGGCTGGGAGAAAGGTGAGCGCATCGAAGAACACCC 1075
QY 321 ArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340
DB 1076 CGCGCGGGGCCAAGAGAGACAAAGTCGCGGATCATCAAGAAAGCACTTCCCTGAGATC 1135
QY 341 ArgLysGlnArgGluLeuGlnGlnLysArgMetGlnSerArgValGlyGlnArgGlySerGly 360
DB 1136 CGCAAGCAGCGCAGCTGCAGAGCGCATGCAGACAGAGGGTGGGCCAGCGGGCAGTGGG 1195
QY 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
DB 1196 CTGTCCATGTCCGCCCGCCGACAGCAGAGAGTGTCAAGATCATCATGTGCTCTCA 1255
QY 381 GluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProPheMetLeuTyr 400
DB 1256 GAGCAGAGAACTTGAGAAAGACAGATCCGCCACTGGCCCTGATCCGCCCATGCTTAC 1315
QY 401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420
DB 1316 GACGCTGACAGCAGCGCATCAAGTTCATCAATGAACGGGCTTATGCGGACCCCATG 1375
QY 421 LysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluYlySerGluThrPhe 440
DB 1376 AAGGTGTACAAACACCGCAGCATGAACATGTGAGTGAAGAGAGAAAGGAGACCTTC 1435
QY 441 ArgGluLysPheMetGlnHisProLysAsnPheGluLeuIleAlaSerPheLeuGluArg 460
DB 1436 CGGGAAGATTCATGCAGCATCCCAAGAACTTTGGCTGATCCCATCTCTCGAGAGG 1495
QY 461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480
DB 1496 AAGACAGTGGCTGAGGCGTCTCTTACTTACTGACTGAAGAAATGAGAACTATTAAG 1555
QY 481 SerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGln 500
DB 1556 AGCCTGTGAGAGGAGCTATCGGCGCGCGGCAAGACAGCAACAAAGCAGCAG 1615
QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluLys 520
DB 1616 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1675
QY 521 AspGluLysGluLysGluLysGluLysGluLysGluLysGluLysProGluValGluAsn 540
DB 1676 GATGAG 1735
QY 541 AspLysGluAspLeuLeuLysGluLysThrAspAspThrSerLysGluAspAspGlu 560
DB 1736 GACAAAGAAACCTCTCAAGAGAAACAGACGAACTCCAGGGAGGAGCAACGAG 1795
QY 561 LysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGly 580
DB 1796 AAGAGGCTGTGCTTCCAAAGCCCGCAAACTGCCAAACAGCCAGGAGAGACCAAGGC 1855
QY 581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600
DB 1856 CGCATCACCGCGCTCATGTGCTATGAGGCCCAACAGAGAGAGGCGCATCAACCCCGCAG 1915
QY 601 SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMet 620
DB 1916 AGCGCGAGCTGCTCCATGAGCTGAATGAGATTCTCGCTGACAGAGAAAGAAATG 1975

QY	621	GIuThrAlaIysIysGlyIleuLeuGlnHieGlyArgmentProSerAlaIleAlaArgMet	640
Db	1976	GAACACACCAAGAAAGGTCTCTGZAACACGGCCGCAACTGTGTGGCCATGCGCCGATG	2035
QY	641	ValGIserIysThrValSerGlnCysIysAsnPhetYrPheAsnYrIYsIysArgGln	660
Db	2036	GTGGGCTTCAAGACTGTGTGTCCCACTGTAGAACTTACTTCACTCACTCAAGAAAGCGAG	2095
QY	661	AsnLeuAspGluIleLeuGlnGlnHieIleValLeuIysMetGluYsgIuaArgAsnAlaArg	680
Db	2096	AACTTCATGATGATCTTGCACACACACAACTGAAATGTGAGAAAGGAGAAAGCAACGGCGG	2155
QY	681	ArgIysIysIysIysAlaProAlaAlaIaIaSerGluGluIuaIaAlaPheProProValIal	700
Db	2156	AGGAAGAAAGAAAGCGCCGGCGAGCGCCAGCAGAGAGGGCTGTGATTCGCCCGGTGTG	2215
QY	701	GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGIuGlu	720
Db	2216	GAGATATAGAGATATGAGAGCCTCGGGCCTGACCGGAATGAGAGAGATGTGTGAGAGAG	2275
QY	721	AlaGluAlaIeunHieAlaIaSerGlyIAsnGluValProArgGlyIuCYeSerGlyProAla	740
Db	2276	GCTGAACCTTACATGCTCTGTGGAAATGAGGTGCCAGAGGGAGAAATGAGTGGCCAGCC	2335
QY	741	ThrValAsnAsnSerSerAspThrGluSerIleProSerProHietHrGluAlaAlaIys	760
Db	2336	ACTGTCAACAACACTCAGACACGAGAGCATCCCTCTCTTCACTCACTGAGCGCCGCAAG	2395
QY	761	AspThrGlyGluAsnGlyProIysProProAlaThrIleuGlyAlaAspGlyProProPro	780
Db	2396	GACACAGGGCAGAAATGGGCCCAAGCCCCCAGCCACCTTGGGGCGCCAGCGGCCACCCCA	2455
QY	781	GlyProProThrProProProArgArgThrSerArgAlaProIleGluProThrProAlaSe	800
Db	2456	GGGCACCCACCCACCCACACCGAGAGAAATCCCGGC-CCCACTAGGCCACCCCGGCTCT	2514
QY	800	rgIuaIaThrGlyAlaProThrProProProAlaProProSerProSerAlaProProPro	820
Db	2515	TGAAGCCACCGGAGCCCTTACGCCCCCAACGACCCCATTCGCTCTGTGACCTCTCC	2574
QY	820	ovaIValProIysGluGluIuYsgIuGluGluThrAlaAlaAlaProProValGIuGluGI	840
Db	2575	TGTGTCTCCCAAGAGAGAGAGAGAGAGAGAAACCGACGACGCCCCCACTGAGAGAGAGG	2634
QY	840	YGIuGIuGluIuYsProProAlaAlaGIuGluIleuAlaValAspThrGlyIysAlaGIuGI	860
Db	2635	GGAGAGAGAGAAGCCCCCGCGCGGTGAGAGACTGTGACTGACACAGGGAAGCGCGAGA	2694
QY	860	uProValIysSerGluCYsThrGluGluAlaGIuGluGlyProAlaIysGlyIysAspAl	880
Db	2695	GCCCGTCAAGAGCGAGTGCAGCGAGGAAGCCGAGAGGGCGCGCCCAAGGGCCAAAGCA	2754
QY	880	agIuaIaAlaGIuAlaThrAlaGIuGIuGlyAlaIleuIysAlaGIuIuYsIysGluYGIYsE	900
Db	2755	GGAGGCGCTGAGGACACGCGCGAGGGGGCCCTCAAGCAAGAAAGAGAGGGCGGGAG	2814
QY	900	rgIYArgAlaThrThrAlaIysSerSerGlyAlaProGlnAspSerAspSerSerAlaTh	920
Db	2815	CGGAGAGGGCCACCAAGCCAAAGACTCGGGGGCCCCCCAGACAGCACTCAATGTCTAC	2874
QY	920	rCYeSerAlaAspGluValaAspGluAlaGIuGIuGlyIYAspIysAsnArgIeuleuSerPr	940
Db	2875	CTGCAGTGCAGACAGAGGTGATGAGAGCGGAGGGCGGCGACAAGAACCGGCTGTGTCCC	2934
QY	940	oArgProSerIleuIeunHrProThrGlyIAspProArgAlaAsnAlaSerProGlnIYsPr	960
Db	2935	AAAGCCACAGCTCTTCAACCCGACCTGGGAGACCCCGGGCCAAATGCTTCAACCCAGAA	2994
QY	960	oLeuAspIeunIYsgIuIeunIYsgIuArgAlaAlaAlaIleProProIleGluValThrIY	980
Db	2995	ACTGGAAGCTTGAAGAGCTGAGAGCAGCGAGCGGCTGTCCATCCCCCATTCAGGTCA	3054
QY	980	sValHieGIuProProArgGluAspAlaAlaProThrIysProAlaProProAlaProPr	1000

Db	3055	AGTCATTAAGCCCCCGGGAGGAGCAGACTCCACCAAGACAGTCCCCCGAGCCACACC	3114
Qy	1000	oProBiogInanLeuGInProGluSerAspAlaProGInGInProGlySerSerProAr	1020
Db	3115	GCCACCGCAAAACCTGACAGCCGAGAGCGACCCCTCGACAGCCGCGAGCGCCCG	3174
Qy	1020	gGlyYsSerArGserProAlaProProAlaAspIysGlyAlaPhaAlaAlaGlyAlaG	1040
Db	3175	GCGCAAGAGCAGAGACCCCGGACCCCCCGCCACAAAGAGGGCTTCGACGCCAGGCCA	3234
Qy	1040	nLySerProGlyAspProProCySerTrpHisSerGlyLeuAspProPheProValProProAr	1060
Db	3235	GAGCTGGCTGGGGAGCCCCCTGTGAGACTTCGGCGCTGCCCTTCCCGTGGCCCCCG	3294
Qy	1060	gGluValIleIysAlaSerProHisAlaProAspProSerAlaPheSerTyAlaProPr	1080
Db	3295	TGAGGTGATCAAGGCTCTCCCGCATGCCCCCGGACCCCTCAGCGCTTCTCTACGCTCACCC	3354
Qy	1080	oGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProPr	1100
Db	3355	TGGTCACCACTGCCCCCGGGCTCATAGACCTGCGCGCGTCTCGCCGCGCCACC	3414
Qy	1100	cThrIleSerAsnProProProLeuIleSerSerAlaIysHisProSerValLeuGluAr	1120
Db	3415	CACCATCTCCAAACCGGCTCTCCCTCATCTCTCTGCGCAAGCACCCAGCGTCTCGAGAG	3474
Qy	1120	gGlnIleGlyAlaIleSerGInGlyMetSerValGlnLeuHisValProTyrSerGluHis	1140
Db	3475	GCAATTAAGTGCATCTCCCAAGAGATGTCGATCCAGCTCCACGTCCGTACTCAGAGCA	3534
Qy	1140	sAlaIysAlaProValIglyProValThMetGlyLeuProLeuProMetAspProIysIy	1160
Db	3535	TGCCAAGGCCCGCGTGGCGCTGTGACCATGGGGCTGCCCCCATGAGACCCCAAAA	3594
Qy	1160	sLeuAlaProPheSerGlyValIysGlnGlnLeuSerProAsxGlyGlnAlaGlyPr	1180
Db	3595	GCTGCACCCCTTCACGGAGGTAGAGAGAGAGAGCTGTCCCACGGGGCCAGGCTGGGCC	3654
Qy	1180	oProGluSerLeuGlyValProThrAlaGlnGluAsaSerValLeuArgIyThrAlaIe	1200
Db	3655	ACCGGAGAGCGCTGGGGGGTCCACAGCCCAAGAGGGTTCGATGTAGAGGAGCAGCTCT	3714
Qy	1200	uGlySerValProIglyIySerIleThrIyGlyIleProSerThrArgValProSerAs	1220
Db	3715	GGGCTCAGATTCCGGCGCGAGCATCACCAAGGCAATCCACAGACACCGGGTGCCTCGGA	3774
Qy	1220	pSerAlaIleThrTyrArgGlySerIlePheHisGlyThrProAlaAspValLeuTyrIy	1240
Db	3775	CAGCCCATCAATACCGGGCTCATCAACCAAGGACGACCGAGCTGAGCTCTGTACAA	3834
Qy	1240	sGlyThrIleThrArgIleIleGlyIyAspSerProSerArgLeuAspArgIyArgG	1260
Db	3835	GGGCACCATCACAGAGATCATGGCGAGAGACGCCGAGTCCGCTTGAGACCGCGCGGGA	3894
Qy	1260	uAspSerLeuProIysGlyHisValIleTyGlnGlyIySlyGlyHisValLeuSerTy	1280
Db	3895	GGACAGCTCGCCCAAGGGCGACATCATCTACGAAGGCAAGAGGGCCACAGTCTTGCTTA	3954
Qy	1280	rGlnGlyIyGlyMetSerValThrGlnCySerIyGlnIyAspGlyIyArgSerSerSerIyPr	1300
Db	3955	TGAGGGTGGACATGTCTGTGACCAAGTCTCCAAAGGACGGCAGAAAGAGCTCAAGAAC	4014
Qy	1300	oProHisGluThrAlaAlaProIyArgThrTyrIyAspMetGlnGlyIyArgValGlyAr	1320
Db	4015	CCCCCATAGACGGCGCCGCCCAAGGACCTTAATGACATGATGAGGGCGCGCTGGGGCAG	4074
Qy	1320	gAlaIleSerSerAlaSerIleGlnGlyIyLeuMetGlyIyArgAlaIleProProGluArgHis	1340
Db	4075	AGGCATCTCTCAGCAGCATGAAAGGTCTCATGGCGCGTGCCATCCCGCGGAGCGACA	4134
Qy	1340	sSerProHisHisLeuIyGlnGlnHisIleIleArgIySerIleThrGlnGlyIlePr	1360

Db 4135 CAGCCCCACCACTCAAGAAGACGACCAATCCGGGGCTCATCAACAAGGAATCCC 4194
Qy 1360 oATSerTyrValGluAlaGlnGluAspTyrLeuArgGluAlaIleuLeuLeuAsp 1380
Db 4195 TCGGTCTTAAGTGAAGGACAGAGACTACCTCGTCGGAGGCCAAGCTCTTAAAGCG 4254
Qy 1380 gGJGJGJYThProProProProProSerArgAspLeuThrGluAlaTyrIleThrG 1400
Db 4255 GAGAGGACGACCTCGCCCCACCGCCCTCACTGAGACCTGACCGAGGCTTCAAGAGCA 4314
Qy 1400 nAlaIleuGlyProleuLeuLeuAspProAlaIleGluGlyLeuValAlaThrValIly 1420
Db 4315 GGGCTTGAGGCCCTTGAAGCTGAAGCGCGCCATGAGAGGCTGTGGCCACAGTGAAGGA 4374
Qy 1420 uAlaGlyArgSerIleHisGluIleProArgGluGluIleuArgHisThrProGluLeu 1440
Db 4375 GGGGGGGCGCTCATCATGAGATCCCGCGAGAGACTCGGACACAGCCCGAGCTGCC 4434
Qy 1440 oLeuAlaProArgProleuLeuIlySerIleThrGlnGlyThrProleuLeuTyrAs 1460
Db 4435 CTTGGCCCCCGCGCGCTCAAGAGAGGCTCATCAACGAGGACCCCGCTCAAGTACGA 4494
Qy 1460 pThrGlyAlaSerThrThrGlySerIlyIyHisAspValArgSerIleGlySerPr 1480
Db 4495 CACCGGGCGCTCCACCACTGGCTTCAAAAGCAGAGCTACGCTCCCTCATCGGCACGCC 4554
Qy 1480 oGlyArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeu 1500
Db 4555 CGGGCGGACCTTCCACCCGCTGACCCGCTGAGTGTATGAGCCGACCCCGGGGACTGGA 4614
Qy 1500 uArgAlaCysTyrGluGluSerLeuIySerArgProGlyThzAlaSerSerSerGly 1520
Db 4615 ACGGCTGCTTACGAGAGAGCGCTGAAGACCGGCGCAGGACCGCCGACGACCTCGGGGG 4674
Qy 1520 ySerIleAlaArgGlyAlaProValIleValProGluLeuGlyIyAspProArgGlnSer 1540
Db 4675 CTCATTGGCGCGCGCGCCCGCTCATTTGCTTGACCTGGACCGCGCGCGCGAGACCC 4734
Qy 1540 oLeuThrTyrGluAspHisGlyAlaProPheAlaGluHisLeuProArgGlySerPro 1560
Db 4735 CTTAACTTATGAGGACCAAGGGGACCTTTGGCGGACCTCCCAAGAGTTGCGCGCT 4794
Qy 1560 lThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSerIyAlaSe 1580
Db 4795 GACCAAGCGGAGCGCCACCGCGCGCTGACGAGGAGCGCTTTGCTTCCAGCAAGCATC 4854
Qy 1580 rGlnAspArgIyLeuLeuThrSerThrProArgGluIleAlaIySerProHisSerThrVa 1600
Db 4855 CCAAGACCGGAAGCTGACGTGACGCGCTCGTGAATGCGCAAGTCCCGGACAGCACCGT 4914
Qy 1600 lProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyValSer 1620
Db 4915 GCGCGAGACACCAACCAACCTCTGCGCTTGAACACCTGCTTGCGGGCGGAGTGG 4974
Qy 1620 yValaAspLeuTyrArgSerHisIleProleuAlaPheAspProThrSerIleProArg 1640
Db 4975 CGTGAACCTGTATGAGGACCAATCCCGCTGCGCTTGAACCACTCCATACCCGCGG 5034
Qy 1640 yIleProLeuAspAlaAlaAlaIaTyrIleuProArgHisLeuAlaProAspPro 1660
Db 5035 CATCCCTCTGAGAGCGCGCTGCTACTACTGCGCCGACACTGCGCCCAACCCAC 5094
Qy 1660 rTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeu 1680
Db 5095 CTAACCGACCTTATCCACCTTATCTCATCCGCGGCTACCCGACAGCGGGCGCTGGA 5154
Qy 1680 uAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisIleAsnThrAl 1700
Db 5155 GAACCGGACAGCATCATCATGCTCATCATCTGCAAGATGACCAACAACGCGGC 5214
Qy 1700 aThrAlaMetAlaGlnArgAlaAspMetLeuArgIyLeuSerProArgGlnSerSerIle 1720
Db 5215 CACCGCATGGCCAGAGACTATATGTGAGGGGGCTCTCGCCCGCAAGTCTCGCT 5274

Qy 1720 uAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisIle 1740
Db 5275 GGCACCTAACTACGCTCGGGGTCCCGAGGACATATGACTGTGCTCCAAAGTCCACACT 5334
Qy 1740 uProValLeuValProProThrProGlyIyThrProAlaThrAlaMetAspArgLeuAla 1760
Db 5335 GCGTGTGCTGTGCTCGCCCGACACAGGACCCCAAGCCGACCTGAGCCGCTTGCTTA 5394
Qy 1760 rLeuProThrAlaProGlnProPheSerArgHisIleSerSerProleuSerProG 1780
Db 5395 CTTCCCAAGCGCGCCCAAGCTTTCAGACAGCCGCAAGAGCTCCCACTTCCAGG 5454
Qy 1780 yGlyProThrHisLeuThrIyAspProThrThrThrSerSerSerGluArgAlaAsp 1800
Db 5455 AGGTCCAAACACTTGAACAACAACCAACCAACAGTCTCTCCGAGGGAGGAGGACCG 5514
Qy 1800 gAspArgGluArgAspArgAspArgGluArgIyIySerIleLeuThrSerThrThr 1820
Db 5515 GATTCGAGAGCGGAGCGGGATCGGAGCGGAAAGTCCATCCTCACTCCACACAGAC 5574
Qy 1820 rValGluHisAlaProIleThrArgProGlyIyThrGlnSerSerGlySerSerGlySe 1840
Db 5575 GGTGAGACAGCACCACTGAGACCTGTATACAGAGACAGAGCGGCGAGCGGAG 5634
Qy 1840 rSerGlyGlyGlyGlyIySerSerSerArgProAlaSerHisIleHisGlnHis 1860
Db 5635 CACGGCGGGGGTGGGGGCGAGCAGCGCCCGCTCCCATCTCCATCCACACAGCA 5694
Qy 1860 sSerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAs 1880
Db 5695 CTCGCCATCTCCCTCGAGCCAGAGATGCCCTCCAGACAGACCCAGTGTGCTTCAAA 5754
Qy 1880 nThrGlyMetIyGlyIleIleThrAlaValGluProSerIyAspProThrValLeuAsp 1900
Db 5755 CACAGCATGAAGGATATCATCACCGCTGAGGCCACAGACCGCACCGTCTCGAGTGC 5814
Qy 1900 rThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCysPr 1920
Db 5815 CACTTCACCTCTTACCCGCTTGCGCGGCTGCGACATTCACCTCCACCACTGCGCC 5874
Qy 1920 oLeuGlyGlyIyThrLeuAspGlyValIyThrProThrLeuMetGluProValLeuLeuPro 1940
Db 5875 ACTGGCGGACCTCGATGGAGGTACCTTACCTCATGAGACCGCTTGTGCGCCAA 5934
Qy 1940 gGluAlaProArgValAlaArgProGluArgProAlaGluAlaAspThrGlyHisAlaPhe 1960
Db 5935 GAGAGCGCCCGGGGTGCGCGGCGAGAGCGCCGAGAGACCGGCGCATCTTCTCT 5994
Qy 1960 uAlaIySPProProAlaArgSerGlyLeuGluProAlaSerSerProSerIySerGly 1980
Db 5995 CGCCAAAGCCCGGCGCGCTCGGGGCTGAGAGCCGCTCTCCCGACAGAGGGCTCGGA 6054
Qy 1980 uProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAla 2000
Db 6055 GCGCGGCGCGCTGCTGCTGCTCTGCGGCGACCATGCGCGGACCGCTGGA 6114
Qy 2000 sAlaLeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSer 2020
Db 6115 GAACCTGACCACTCACAGCGCGCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGG 6174
Qy 2020 pProHisArgGluIyThrGlnSerIySPProPheSerIleGlnGluLeuGluIySer 2040
Db 6175 CCGGACCGGGAAGAAACTCAAGTAAACCTTTTTCATTCAGAACTGAACTCCGTTTC 6234
Qy 2040 rLeuGlyIyThrHisGlySerSerTyrSerProGlnGlyValGluProValSerProVal 2060
Db 6235 TCTGGGTATACACGAGAGAGCTACAGCCCGGAAGGGGTGAGGCCGTGAGCCTGTGAG 6294
Qy 2060 rSerProSerLeuThrHisAspIyGlyLeuProIyHisIleGluGluLeuAspIySse 2080
Db 6295 CTCACCACTGACCAACGACAGCAAGGGGCTCCCAAGACCTGGAAGAGCTGACAAAGAG 6354

QY 2080 rhi bleuglvglyglu leuargProlysglnProglyProvalylleuglyglu 2100
DB 6355 CCACCTGGAGGGGAGAGCTCGGCGCAAGAGCCAGCCCGGTGAACCTTGGCGGAGGCG 6414
QY 2100 aAlaHsleuProHsleuargProleuProgluSerInProSerSerProleu 2120
DB 6415 CGCCCACTCCCAACACTCGGCGGCTGAGACCAAGCCCTCGTCCAGCCGCTGCT 6474
QY 2120 uclnThrAlaProglyValylleuglyHsleuargValAlThrleuAlaglHsleu 2140
DB 6475 CCAGACCGGCGCCAGGGGTCAAGGTACCAAGCGGGGTGTCACCTGGCCAGCACTCAG 6534
QY 2140 rgluValAlIethrAlaSPTrThrargHsHsleuProglInleuSerAlaProleuP 2160
DB 6535 TAGAGTCATCACACAGACTACACCCGAGCACCCACAGACACTGAGCGCACCCCTGCC 6594
QY 2160 oAlaPleuTyrrSerPheProglValaserySProvalleuAPleuargArpPro 2180
DB 6595 CGCCCTCTACTCTCTCTGCGGCGAGCTGCCCGTCTGACCTCGCGCCGACCC 6654
QY 2180 oSerAPleuTyrrleuProProProAspHsleuAlaProAlaArgGlySerProHs 2200
DB 6655 CAGTGACCTCTACTCTCCGCGCGGACCATGTGCCCGCGCGCTCCGCGCCACAG 6714
QY 2200 rgluGlyglYlySPargSerProgluProAsnlySPHrSerValleuGlyglYgluAs 2220
DB 6715 CCAAGGGGCGCAAGGCTCTCCAGAGCCAAACAGAGCTCGTCTTGAGTGTGTAGGA 6774
QY 2220 pglYlIegluProvalSerProProgluGlyMetThrInuProglYHsleuArgSerAl 2240
DB 6775 CGGTATGAACTGTGTCCCGGAGGAGCATGAGAGCCAGGCGACTCCCGAGTGC 6834
QY 2240 aValTyrrleuLeuTyrrArgSPglYgluInThrInuProSerArgMetGlySerly 2260
DB 6835 TGTGTACCGGCTCTCTACCGGAGTGGGAGACAGAGGAGCCAGAGATGGCTCCAA 6894
QY 2260 sSerProGlyAsnThrSerGlnProProAlaPhePheSerlyleuThrGluSerAnse 2280
DB 6895 GTCTCCAGGCAACACAGCAGCGCGAGCTTCTCAGCAACCTGACCCAGAGCAACTC 6954
QY 2280 rAlaMetVallySerlyleuGlnIuIeAsnlySlyleuAsnThrHsAsnArgAs 2300
DB 6955 CGCATGTCTCAAGTCAAGAGAGATCAACAGAGCTGTAACCCACAAACCGGAA 7014
QY 2300 ngluProGluTyrrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIeth 2320
DB 7015 TGAAGCTGAATCAATATCAAGCCAGCTGAGAGAGATTTCAATATGCCCGCCATCAC 7074
QY 2320 rglYThrGlyleuMetThrTyrrArgSerGlnAlaValGlnGluHsleuAsnThrAnMe 2340
DB 7075 CGGAACAGGCTTATGACCTATGAAGCCAGCGGTGCAAGAACATGCCAGCAACAT 7134
QY 2340 rglYleuGluAlaIleIeArglySlyleuMetGlylySPHrArgSPglInProgluIuse 2360
DB 7135 GGGGCTGAGAGCCATATTAAGAAAGCATCATGGTAAATATGACCAAGGGGAAAGTCT 7194
QY 2360 rProProleuSerAlaAsnAlaPheAsnProleuAsnAlaSerAlaSerleuProAlaAl 2380
DB 7195 CCGCGCGCTCAGGCGCAATGCTTTAACTCTGAATGCCAGTCCAGCTGCCCTGCTGC 7254
QY 2380 aMetProIleThrAlaAlaAspGlyArgSerAspHsleuThrSerProglYglYgl 2400
DB 7255 TATGCCCATTAACCGCTGTGACGAGAGGTGACCAACACTCAGCTCCGCAAGTGGCG 7314
QY 2400 yglYlySPAlaIlySPValSerGlyArgProSerSerArglyValAllySPProAlaProgl 2420
DB 7315 CGGAAAGGCGAAAGTCTCTGGACAGCCAGAGCCAGAAAGCCAAAGTCCCGCGCGG 7374
QY 2420 yleuAlaSerGlyAspArgProProSerValIleSerSerValHsleuGlyglYasPCyAs 2440
DB 7375 CCGGAGCATCTGGGAGACCGGCAACCTGTCTCTCAGTCACTCGAGGAGAACTGCAA 7434
QY 2440 nArgArgThrProleuThrAsnArgValTrpGluAspArgProSerSerAlaGlySerTh 2460

DB 7435 CCGCGGAGCGGCTCACCAACCGGTGTGGAGAGACAGGCTGTCCGACAGTTCAC 7494
QY 2460 rProPheProTyrrAsnProleuIleMetArgleuGlnAlaGlyValMetAlaSerPro 2480
DB 7495 GCCATTCCCTTCAACACCCCTGATCATGCGTGGACAGGGGTGTGATGCTTCCCAAC 7554
QY 2480 oProProGlyleuProAlaGlySerGlyProleuAlaGlyProHsleuAlaTrpAsp 2500
DB 7555 CCACCGGCGCTCCCGCGGCGAGCGGCGCTCGCTGCGCCCGCACAGCCTGGGACGA 7614
QY 2500 uGluProlySPleuLeuCySerGlnThrGluThrleuSerAspSerGlu 2517
DB 7615 GAGGCCAAGCCACTGTGCTGCGATGACAGACACTTCGACAGCGAG 7666

RESULT 5

US-09-819-104A-3
; Sequence 3, Application US/09819104A
; Publication No. US20030027137A1
; GENERAL INFORMATION:
; APPLICANT: Chen, J. Don
; TITLE OF INVENTION: NOVEL, NUCLEAR RECEPTOR COREPRESSOR MOLECULES
; TITLE OF INVENTION: AND USES THEREFOR
; FILE REFERENCE: UMG-030
; CURRENT APPLICATION NUMBER: US/09/819,104A
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/193,138
; PRIOR FILING DATE: 2000-03-29
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 7521
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(7521)
US-09-819-104A-3

Alignment Scores:

Pred. No.: 0 Length: 7521
Score: 12978.00 Matches: 2487
Percent Similarity: 98.578 Conservative: 3
Best Local Similarity: 98.468 Mismatches: 9
Query Match: 98.214 Indels: 28
DB: 10 Gaps: 3

US-09-522-753-5 (1-2517) x US-09-819-104A-3 (1-7521)

QY 1 MetSerGlySerThrGlnleuValAlaGlnThrTPArgAlaThrGluProArgTyrrPro 20
DB 1 ATGTCCGGGCTCCACACAGCCTGTGGCAGAGGTGAGGCGGCATGAGCCCGCTACCGG 60
QY 21 ProHsleuSerleuSerTyrrProValGlnIleAlaArgThrHsleuAspValGlyleu 40
DB 61 CCCACAGGCTTTCTTACCCAGTGCAGATGCGCGGAGCAACAGAGCTGGGCTCTTG 120
QY 41 GluTyrglnHsleuSerArgSPTrAlaSerHsleuSerProglYserIleIegin 60
DB 121 GAGTACACACACACTCCCGGACATATGCTCCACCTTCCCGGCTCCATCATCCAG 180
QY 61 ProGlnArgArgProSerleuLeuSerGluPheGlnProglYAsnGluArgSerGln 80
DB 181 CCCAGCGGCGAGGCGCTCCCTGCTGTGAGTTCCAGCCCGGGAATGAACGCTCCAG 240
QY 81 GluLeuHsleuArgProgluSerHsleuTyrrleuProgluIuIylySPSerGluMet 100
DB 241 GAGCTTCACTCGGCGAGAGTCCCACTATTCCTGCGAGCTGGAGATCAAGATG 300
QY 101 GluPheIleGluSerlySPArgProArgleuGluLeuLeuProAspProleuLeuArgPro 120
DB 301 GAGTTCAATTGAAGCAAGCGCCTCGGCTAAGAGCTGCTCTGACCCCGTGTGCGAACCG 360

QY 121 SerProLeuLeuValaThrglyGlnProAlaGlySerGluAspLeuThrLyAspArgSer 140
Db 361 TCACCCCTGGTGGCCAGCGGCGACCTGGCGGATCTGAAGACCTCACCAAGACCGTAGC 420
QY 141 LeuThrLysIlybLeuGluProValSerProPProserProProhiIsthrAspProGluLeu 160
Db 421 CTGACGGGCAAGCTGGAAACCGGTGTCTCCCGGACCGCCCGGCACTGACCTCGAACCTG 480
QY 161 GluleuValProProAlaSerLeuSerLyGluGluLeuLeuLeuLeuLeuLeuLeuLeuLeu 180
Db 481 GAGCTGGTGGCCGACCGGCTCTCCAAAGAGAGCTGATCCAGAACATGAGACCGGCTGAC 540
QY 181 ArgGluIleThrMetValGluGlnGlnIleSerLybLeuLybLybGluGlnGlnLeu 200
Db 541 CGAGAGATCACCATGATGAGACAGACATCTTAAGCTGAAGAAAGAACAGCAACACTG 600
QY 201 GluGluGluValAlaIlybProProGluProGluLybProValSerProProProIleGlu 220
Db 601 GAGAGAGAGCTGGCCAAAGCGGCGGACCTGAGAGAGCGGTGTACCGCGCCCATCGAG 660
QY 221 SerLybIleArgSerLeuValGlnIleIleTyraAspGluAsnArgLybValAlaGluAla 240
Db 661 TCAGAACCGGACCGCTGGTGGACATCTACAGAGAAACGGAAGAGAGCTGAAGCT 720
QY 241 AlaIleArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyraAsnGlnPro 260
Db 721 GCACATCGAATTCTGGAAGGCTGGGGCCCGAGGTGAGGCTGGCGCTGTACACCAACCC 780
QY 261 SerAspThrArgGlnTyraIleGluAsnIleLybIleAsnGlnAlaMetArgLybValLeu 280
Db 781 TCCGACACCGGCGAGTATCATGAGAACATCAAAATAAACAGCGGATCGGAGAAAGACTA 840
QY 281 IleLeuTyraPheLybArgArgAsnHisAlaArgLybGlnTrpLybGlnLybPheCysGln 300
Db 841 ATCTTGACTTCAAGAGAGAGAAATCACGCTCGAAACAAATGGAGAGAGAGTTCTGCGAG 900
QY 301 ArgTyraAspGlnLeuMetGluAlaLeuGluLybLybValGluArgIleGluAsnAsnPro 320
Db 901 CGCTAAGACCAAGCTCATGAGGCTGGGAGAAAGAGTGAAGGCGCATCGAAGCAACCCC 960
QY 321 ArgArgArgAlaLybGlySerLybValArgGluTyraGluLybGluPheProGluIle 340
Db 961 CGGCGGCGGCGCAAGAGAGCAAGTTCGCGAGTACTACAGAAAGCGATGTCCTCGAGATC 1020
QY 341 ArgLybGlnArgGluLeuGlnGlnIleArgMetGlnSerArgValGlyGlnArgGlySerGly 360
Db 1021 CGAAGCAGGCGGAGCTGCGAGAGCGCATCG--AGGCTGGGCGGCGGAGAGTGGG 1077
QY 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
Db 1078 CTGTTCATGTGGCGCGCGCGCGCAGCAGCAAGAGTGTCAAGATCANTCGATGGCTCTCA 1137
QY 381 GluGlnGluAsnLeuGluLybGlnMetArgGlnLeuAlaValIleProProMetLeuTyra 400
Db 1138 GAGCAGAGAGAACTCGAGAGAGATGCGCGCACTGGCGCTGATCCCGCCCATGCTGTAC 1197
QY 401 AspAlaAspGlnGlnArgIleLybPheIleAsnMetArgLybLeuMetAlaAspProMet 420
Db 1198 GAGCGTGAAGAGGCGCATCAAGTTCATCAACATGAACGGGCTTAAGCGGACCCCATG 1257
QY 421 LybValTyraAspArgGlnValMetAsnMetTrpSerGluGlnGluLybGluThrPhe 440
Db 1258 AAGGTGTACAAACACCGCCAGGTCAATGAATGTGAGTGAAGAGAGAGAGACCTTC 1317
QY 441 ArgGluLybPheMetGlnHisProLybAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
Db 1318 CGGGAAGAGTTCATGACATCCCAAGAACTTGGCTGATCCCATCATCTCGAGAGAGG 1377
QY 461 LybThrValAlaGluCysValLeuTyraTyraTyraLeuThrLybLybAsnGluAsnTyra 480
Db 1378 AAGACAGTGGCTAGTGTCTCTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 1437
QY 481 SerLeuValArgArgSerTyraArgArgGlyLybSerGlnGlnGlnGlnGlnGln 500

Db 1438 AGCTGTGAGACGAGACTATCGCGCGCGGCAAGAGCCAGCAGCAACACAGCAGAG 1497
QY 501 Gln 520
Db 1498 CAGCAG 1557
QY 521 AspGluLybGluLybGluLybGluAlaGluLybGluGluLybGluLybGluLybGluLyb 540
Db 1558 GATGAG 1617
QY 541 AspLybGluAspLeuLeuLybGluLybThrAspAspThrSerGlyGluAspAsnAspGlu 560
Db 1618 GACAAAGAGACCTCTCAAG 1677
QY 561 LybGluAlaValAlaSerLybGlyArgLybThrAlaAsnSerGlnGlyArgGlyGly 580
Db 1678 AAGAGGCTGTGGCTTCCAAAGGCGCCAAATCTGCCAAACAGCAGGAGAAACCGCAA 1737
QY 581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600
Db 1738 CGCATCACCGCTCATGCTTAATGAGGCCAAAGCAGAGAGAGAGAGAGAGAGAGAGAG 1797
QY 601 SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluMet 620
Db 1798 AGCGCGAGCTGGCTTCATGAGAGCTGAATGAGAGTCTGCTGGACAGAAAGAAATG 1857
QY 621 GluThrAlaLybLybGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMet 640
Db 1858 GAAACAGCCCAAGAAAGTCTCTGGAAACAGCGCGCAACTGTGCGGCATCGCCGATG 1917
QY 641 ValGlySerLybThrValSerGlnCysLybAsnPheAsnTrpPheAsnTrpLybValArgGln 660
Db 1918 GTGGGCTCCAGACTGTGTGCGAGTGAATGAACTTCTTAATCACTACAGAAAGAGCAG 1977
QY 661 AsnLeuAspGluIleLeuGlnGlnHisLybLeuLybMetGluLybGluArgAsnAlaArg 680
Db 1978 AACCTCATGATGATCTTGGAGCAGCAGCAAGCTGAAGTGAAGAGAGAGAGAGAGAG 2037
QY 681 ArgLybLybLybValAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValAla 700
Db 2038 AGGAAAGAGAAAGAAACCGCGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAG 2097
QY 701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGlu 720
Db 2098 GAGCATGAGAGATGAGAGGCTGCGCGTGAACCGAAATGAGAGAGATGTGAGAGAG 2157
QY 721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAla 740
Db 2158 GCTGAA-----GCC 2166
QY 741 ThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLyb 760
Db 2167 ACTGTCAACAAACGCTTCAGACACCGAGAGATCCCTCTCTCAACTGAGGCGCGCAAG 2226
QY 761 AspThrGlyGluAsnGlyProLybProProAlaThrLeuGlyAlaAspGlyProProPro 780
Db 2227 GACACAGGAGAGATGGGCCCAAGGCCCAAGCCCACTGGGCGCGAGAGAGAGAGAGAGAG 2286
QY 781 GlyProProThrProPro-ArgArgThrSerArgAlaProIleGluProThrProAla 800
Db 2287 GGGCCACCAACCCCAACCGAGAGAGATCCGGGC-CCACAGATGTCCACCCCGGCTC 2345
QY 800 TGTuAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProProPr 820
Db 2346 TGAAGCACACTTAAGCCCTACAGCCCGCCACAGACCCCATTTCCCTTCACTCTCC 2405
QY 820 oValValProLybGluGluLybGluGluGluThrAlaAlaAlaProProValGluGlu 840
Db 2406 TGTGTGCTCCCAAG 2465
QY 840 yGluGluGluLybProProAlaAlaGluGluLeuAlaValAspThrGlyLybAlaGluGlu 860

Dh 2466 GGAGAGCAGAAAGCCCCCGGCTGAGGAGCTGCGACTGTGACACAGAGGAGGCCGAGGA 2525
Qy uProVallyserGluCysThrGluGluValIaGluGluGluYProAlaIyGluYLeuAspAl 880
Dh 2526 GCCCGTCAAGAGCGATGCAAGGAGGAGCCGAGAGGGCCGAGCCCAAGGAGCAAGAGCCG 2585
Qy 880 aGluValaIaGluValaThrAlaGluValaLeuYValaGluYValaGluYValaGluYVala 900
Dh 2586 GGAGGGCCGCTGAGAGCCAGGAGCGGCTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 2645
Qy rGluYValaIaThrThrAlaIySerSerGluYAlaProGluAspSerSerAlaTh 920
Dh 2646 CGGAGAGGGCCAGCCAGCCAGAGCTCGGAGCGCCCGCCAGAGCAGGAGCTCCAGTCCAC 2705
Qy rCysSerAlaAspGluValaAspGluValaGluGluYValaAspYAspAlaArgLeuLeuSerPr 940
Dh 2706 CTGCACTGAGAGAGAGTGAATGAGCCAGAGGGCGGAGACAGAACCGGGCTGCTGTCCCC 2765
Qy oArgProSerLeuLeuThrProThrGluYAspProAlaAlaAspAlaSerProGluYLeuSerPr 960
Dh 2766 AAGGCCAGACCTCTCCACCCCGACTGGCGACCCCGGAGCAATGCTCCACCCAGAGAGCC 2825
Qy oLeuAspLeuYValaGluYValaGluYValaAlaAlaAlaIleProProGluIleGluValaThrY 980
Dh 2826 ACTGAGCTTGAAGCAGCTGAAGCAGCAGCGCTGCTCATCCCTCCATCCAGGCTCACCAA 2885
Qy sValHisGluProProArgGluValaAspAlaAlaProThrYAspProAlaProProAlaProPr 1000
Dh 2886 AGTCCATGAGCCCCCGGAGAGAGCCAGCTCCACCAAGCCAGCTCCCGAGCCGAGCC 2945
Qy 1000 oProProGluAspLeuGluProGluSerAspAlaProGluGluProGluYSerSerProAl 1020
Dh 2946 GCCACCGCAAAACCTGACAGCGGAGAGCAGCGCCCTCAGACAGCTGAGCAGCGCCCG 3005
Qy 1020 gGluYLeuSerArgSerProAlaProProAlaAspYLeuY----- 1033
Dh 3006 GGGCAGAGAGCAGAGCGCCGAGCCCGCCGAGCAGAGAGGAGAGAGCCGCTGTCTT 3065
Qy 1034 ----AlaPheAlaIaGluValaGluYLeuYProGluYAspProProCysThrThrSerG 1052
Dh 3066 CCCAGGCTTTCAGAGCCGAGAGCCAGAAAGCTGCTGGAGAGCCCTTGTGCTGAGCTTCCGG 3125
Qy 1052 YLeuProPheProValProProAlaGluValaIleYValaSerProHisAlaProAspPr 1072
Dh 3126 CCGAGCCCTTCCCGTCCCGCCCGGTGAGGTGATCAAGGCTCCCGCATGCCCGAGACC 3185
Qy 1072 oSerAlaPheSerTyAlaProProGluYHisProLeuProLeuYLeuHisAspThrAl 1092
Dh 3186 CTCAAGCTTCTCTTACGCTTCACTGATCACTGATGCCCTGAGCTTCATGACACTGC 3245
Qy 1092 aArgProValLeuProArgProProThrXLeuSerAspProProProLeuYLeuSerAl 1112
Dh 3246 CCGGCGGCTCTCCGCGCCAGCCACCATCTCCAAACCGGCTCCCTCATCTCTCTGC 3305
Qy 1112 alybHisProSerValLeuGluValaGluYValaIleSerGluYLeuSerValG 1132
Dh 3306 CAAGCAGCCAGAGCTCTCGAGAGGCAAAATAGTGCATCTCCCAAGAAATGCGGTCA 3365
Qy 1132 nLeuHisValProTySerGluHisAlaYValaProValGluYProValaThrMetGluY 1152
Dh 3366 GCTCCAGCTCCCTACTCAGAGATGCCAAGGCGCGGTGGGCTGTCTCACTGAGGCT 3425
Qy 1152 uProLeuProMetAspProYValYLeuValaProPheSerGluYValaIleGluYVala 1172
Dh 3426 GCCCTGCTCATGAGCCCAAAAGCTGAGCCCTTCAAGGAGGTGAAGAGAGAGAGCT 3485
Qy 1172 uSerProArgGluYValaGluYProProGluSerLeuYValaProThrAlaGluYVala 1192
Dh 3486 GTCGCCACGGGCGCAGGCTGAGCCACCGAGAGAGCTGGGGGTGCCACAGCCAGAGAGGC 3545
Qy 1192 aSerValLeuArgGluYThrAlaLeuGluYSerValProGluYValaIleThrYLeuYVal 1212
Dh 3546 GTCCGTGCTGAGAGGAGAGCTCTGGGCTCAGTTCCGGGCGAGAGCATCCAAAGGCA 3605

Qy 1212 eProSerThrArgValProSerAspSerAlaIleThrTyArgGluYSerIleThrHisG 1232
Dh 3606 TCCAGAGCAGAGGGTGCCTCGGAGCGGCATCAATCAATCCGGCTCATCACCCAGCCG 3665
Qy 1232 YThrProAlaAspValLeuTyLeuGluYThrIleThrAlaGluYIleGluYValaAspSerPr 1252
Dh 3666 CAGCGCAGCTGAGCTCTCTTCAAGAGGACCATCAACAGAAATCATCGGAGAGAGAGCC 3725
Qy 1252 oSerArgLeuAspArgGluYArgGluYAspSerLeuProYValYHisValaIleTyGluY 1272
Dh 3726 GAGTGCCTTGGAGCCGGCGCGGAGAGACAGCTTCGCCAAGGAGCATCATCAAGAAAG 3785
Qy 1272 YLeuYValaIleValaLeuSerTyGluYValaIleMetSerValaThrGluYSerYVala 1292
Dh 3786 CAAGAAAGGCGCAGCTTCTTCTTCAATGAGGAGGAGCTGTGTGAGCCAGTGTCTCCAAAG 3845
Qy 1292 uAspGluYArgSerSerSerGluYProProHisGluYThrAlaAlaProYValaYThrTyAs 1312
Dh 3846 GAGCGCAGAGAGCAGCTCAGAGACCCCTCATGAGAGAGCGCGCCCAAGCGCAGCTTATGA 3905
Qy 1312 pMetMetGluYArgValaGluYArgAlaIleSerSerAlaSerIleGluYLeuMetG 1332
Dh 3906 CATGATGAGGGCCGCTGGGAGAGCCATCTCTCAGCCAGCATGAGAGCTTCTAGGG 3965
Qy 1332 YArgAlaIleProProGluYValaArgHisSerProHisHisLeuYValaGluYHisIleAla 1352
Dh 3966 CCGTGCATCCCGCGGAGAGCAGACAGCCCACTCAACAAAGAGCAGACACCATCCG 4025
Qy 1352 gGluYSerIleThrGluYIleProArgSerTyValaGluYValaGluYValaAspTyLeuY 1372
Dh 4026 CGGCTCATCAACAAGAGGATCTCTGATCTTACGTGAGGAGCAGAGAGAGACTTCTGGG 4085
Qy 1372 gArgGluYAlaYLeuYLeuYArgGluYValaYThrProProProProProSerArgAs 1392
Dh 4086 TCGGAGAGCCAGAGCTCTTAAAGCGGAGGAGCAAGCTCCGCCCAACCGGCTCAAGGGA 4145
Qy 1392 pLeuThrGluYAlaTyTySerThrGluYValaGluYProLeuYValaYAspProAlaHisG 1412
Dh 4146 CCGAGCCGAGGCTCAAGAGCAGAGGCTCTGGAGCCCTCGAAGCTCAAGCCGCGCATGA 4205
Qy 1412 uGluYLeuValaAlaThrValaYValaGluYValaYArgSerIleHisGluYIleProArgGluY 1432
Dh 4206 GGGCTGTGTGGCAGCGGTAAAGAGGGCGGCTCATCAATGAGATCCCGGAGAGGA 4265
Qy 1432 uLeuArgHisThrProGluYLeuProLeuYAlaProArgProLeuYValaGluYSerIleTh 1452
Dh 4266 GCTGGGAGCAGCGCCGAGACTGCCCTTGCCCGCGCGCTCAAGAGAGGCTCCATTCAG 4325
Qy 1452 rGluYValaProLeuYValaAspThrGluYValaSerThrGluYSerTyValaHisVala 1472
Dh 4326 GAGAGGAGCCCGCTCAAGTACAGACCGGCGGTTCACACATGAGCTCCAAAAGAGAGCA 4385
Qy 1472 pValaArgSerLeuYIleGluYSerProGluYValaArgThrPheProProValaHisProLeuYAsp 1492
Dh 4386 CGTACGCTCCCTCATCTGGAGCGCCGCGAGAGTGTCCACCGGTGACACCCGCTGATGT 4445
Qy 1492 lMetAlaAspAlaArgAlaLeuGluYValaGluYValaCysTyGluYValaSerLeuYSerArgPr 1512
Dh 4446 GATGGCCAGCGCCCGGAGCTGGAAGCTGCTCTGTCAGAGAGAGCTGAAGAGCGGGCC 4505
Qy 1512 oGluYThrAlaSerSerSerGluYValaSerIleAlaArgGluYAlaProValaIleValaProG 1532
Dh 4506 AGGAGACCGCAGAGCTCGGGGGGCTCCATTGCGCGCGCGCGCGGCTGATGTGCTGA 4565
Qy 1532 uLeuYValaYProYValaGluYSerProLeuYThrTyGluYValaPheHisGluYAlaProPheAlaG 1552
Dh 4566 GCTGGGTAAAGCCGCGCAGAGCCCTTGAATGAGAGCAGCGGAGCACTTGTGGCGG 4625
Qy 1552 YHisLeuProArgGluYSerProValaThrMetArgGluYProThrProArgLeuYVala 1572
Dh 4626 CCACTTCCAGAGAGTTCGCCGTGACATGCGGAGGCCACCGCGCGCTGAGAGAGGG 4685

Qy	1572	ySerLeuSerSerSerSerIySaIaSerGlnAspArgIyLeuThrSerThrProAlaGlnI	1532
Db	4686	CAGCCTTTGTCGACGAAGGATCCCAAGACGAAAGCTGACGTGCGCTGTGAAT	4745
Qy	1582	aAlaIySerProHisSerThrValProGlnHisIProHisProIleSerProTyArgI	1612
Db	4746	CGCCAAAGTCCCGACAGACCGTGTCCGAGACCAACCCCACTGTGCGCTTATGA	4805
Qy	1612	uHisLeuLeuArgIyValISerGlyValAspLeuTyraISerSerHisIleProLeuAlaPh	1632
Db	4806	GCACCTGTGGGGCGTGAAGTGCGGTGACCTGTATGCGAGCCACATCCCCCTGGCCTT	4865
Qy	1632	eAspProThrSerIleProAlaGlyIleProLeuAspAlaAlaAlaIyTYrLeuPr	1652
Db	4866	CGACCCCACTCCATACCCCGCGGATCCCTCTGACCGACCCGCTGCTCTACCTGCC	4925
Qy	1652	oArgHisLeuAlaProAsnProThrTYrProHisIleuTyProProTYrLeuIleArgG	1672
Db	4926	CCGACACTGGCCCCCAACCCCACTTACCGCACTGTACCACTCACTCATCTCCGCG	4985
Qy	1672	TYrProAspThraIaIaIeArgIyAsnArgGlnThrIleIleAsnAspTYrIleHnSe	1692
Db	4986	CTACCCCGACAGGGGGGCTGGAGAACCGGCAGACATATCATATGATCATACCTC	5045
Qy	1692	rGlnGlnMetHisIaSnThrAlaIleAlaMetAlaGlnArgAlaAspMetLeuArgI	1712
Db	5046	GCACACAGATGCACCAACAACGGGCACCGCATGGCCACGAGACTGATATGCTGAGGG	5105
Qy	1712	YLeuSerProArgIyIuSerSerIleuAlaIeAsnTYrAlaIaIyProAlaGlyIleI	1732
Db	5106	CCTCTCCGCGCGAGATCTCTGCTGGGACCTCACTACGCTGGGGGTCCCGAGGCATAT	5165
Qy	1732	eAspLeuSerGlnValProHisLeuProValIleuValProProThrProGlyThrProAl	1752
Db	5166	CGACTGTCTCCAAATGGCCACACCTGCTGTCTGTGTCCCGCCGACCAAGCACCCAC	5225
Qy	1752	aThrAlaMetAspArgIeAlaIyTYrLeuProThraIaProGlnProPheSerSerArgH	1772
Db	5226	CACCGCATGAGACGGCTTGCTTACCTCCACCGCGGCCGACCTTACAGACCGGCA	5285
Qy	1772	sSerSerSerProIeuSerProGlyGlyProThrHisIleuThryIyProThrThrTrSe	1792
Db	5286	CAGAGCTCCCGCATCTCCCGAGAGGTCCAAACACTTGCACAAACMACACACACAGTC	5345
Qy	1792	rSerSerGlnArgIyIuArgAspArgAspArgGlnArgAspArgAspArgIuArgIuLy	1812
Db	5346	CTCCTCGAGCGGAGCGACAGACCGGATCGAGACCGGAGTCGGAGCGGGAAAA	5405
Qy	1812	sSerIleuThrSerThrThrThrValGlnHisAlaProIleIleThraIyProGlyThrG	1832
Db	5406	GTCATCTCTACGTCCACACAGACGGGTGGAGACGACCCATCTGGAGACTGTGACGA	5465
Qy	1832	uGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgProAl	1852
Db	5466	GCAAGAGACGGCGCAGCGGACAGACCGCGGGGGGTGGGGGCGACAGCCGCCCGC	5525
Qy	1852	aSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaIeArg	1872
Db	5526	CTCCCACTCCATCCACCAAGCACTGACCTCATCTCCCTCTGGACCAAGATGGCTTCA	5585
Qy	1872	ngIuArgProSerValIeHisAsnThrGlyMetIleGlyIleIleThraIaValGluPr	1892
Db	5586	GCAAGACCCAGTGTGCTTACACACAGGATGAAGGATATATCAACCGTGTGGAGCC	5645
Qy	1892	oSerIyProThryValIleuArgSerThrSerThrSerSerProValArgProAlaAlaTh	1912
Db	5646	CAGCAACCCACGGGTCTGAAGTCCACTCCATCTCTCAACCCCTTGCCAGCTGCAC	5705
Qy	1912	rPheProProAlaThrHisCysProLeuGlnGlyThIleuAspGlyValTYrProThrIe	1932
Db	5706	ATTCCCACTGGCACCCCACTGGCCACTGGGGGGGACCTTGATGGGGGTCTAACCTTACC	5765
Qy	1932	uMeGluProValIleuLeuProIySGlnAlaProArgValAlaArgProGluArgProAr	1952

Db	5766	CATGGAGCCCGCTCTTGCTGCCCAAGAGAGGCCCGCGGCTCGCCGCGCCAGAGCGGCCCG	5825
Oy	1952	GALAAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAl	1972
Db	5826	AGCAGACACCGGCCATGCGCTTCTCTGCCAAGCCCCACCGCGCTCGGGCTGGAGCCCG	5885
Oy	1972	ASerSerProSerLysGlySerGluProArgProLeuValProProAlaSerGlyHisAla	1992
Db	5886	CTCGTCCCCCAGCAAGGGGCTCGAGAGCCCGGCCCTTAGTGCTCTGCTGGCCAGCG	5945
Oy	1992	aThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProPr	2012
Db	5946	CACCATGCGCCGCAACCCCTCGGAAGAACTTGCACTCCACACGCGACCCGAGCCGCC	6005
Oy	2012	oAlaProProAlaSerAlaSerAlaSerAProHisArgGlyuArgThrGlnSerLysProPheSe	2032
Db	6006	GGCGCCACCTGCGCTCGGCTCGAGACCGGACCGGAAAGACTCAAGATGAACCTTTTC	6066
Oy	2032	rIleGlnGluLeuLeuLeuAlaArgSerLeuGlyTyThrIleGlySerSerTyPserProGluG	2052
Db	6066	CATCAGGAACTGGAACTCGGTTCTCTGGGTATCAACGCGAGAGACTACAGCCCGGAAG	6122
Oy	2052	yValGluProValSerProValSerSerProSerLeuThrHisAsnLysGlyLeuProLy	2072
Db	6126	GGTGAGCGCCCTGACGCGCTCGAGACCGGACCGGAAAGACTCAAGATGAACCTTTTC	6185
Oy	2072	hHisLeuGlnGluLeuLeuAsnLysSerHisAlaLeuGluGlyLeuLeuArgProLysGlnProG	2092
Db	6186	GCACCTGGAAAGCTCGCAAGAGCCACTTGAGAGGGAGGCTGGCGGCCCAAGACGCCAGG	6245
Oy	2092	yProValLysLeuLysGlyGluAlaAlaHisAlaLeuProHisAlaLeuArgProLeuProGluSe	2112
Db	6246	CCCCGTGAAGCTTGCGGGGAGAGCGGCCCACTCCACACTGGGGCGCGTGCCTGAAG	6305
Oy	2112	rGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValAlaLysGlyHisGlnAlaGva	2132
Db	6306	CCAGCCCTCGTCCGACCGCGCTGCTCCAGACCGGCCGAGGGGTCAAAAGTCCACAGCGGCT	6365
Oy	2132	lValThrLeuAlaGlnHisIleSerGluValIleThrGlnAsnArgTyThrArgHisHisPr	2152
Db	6366	GGTACACCTGGCCCGACGACATCACTGAGTATATACACAGGACTACACCCGAGCACACCC	6422
Oy	2152	oGlnGlnLeuSerAlaProLeuProAlaProLeuTySerPheProGlyAlaSerCyAPr	2172
Db	6426	ACACGACACTCAGGCGCAACCCCTGGCCGCCCTCTACTCTTCCCTGGGGCAGCTGCC	6485
Oy	2172	oValLeuAspLeuArgArgProProSerAspLeuTyTleuProProAspHisGlyAla	2192
Db	6486	CGTCTCGAAGCTCGCGCGGCCACACCAAGTACTCTACCTCCGCGCCCGGACCATGCGC	6545
Oy	2192	aProAlaArgGlySerProHisSerGlnGlyGlyLysAlaArgSerProGluProAlaLysTh	2212
Db	6546	CCCCGCGCGTGGCTCCCCCAACAGCGAAGGGGAGAAAGGTCTCCAGAGCCAAACAAAC	6605
Oy	2212	rSerValLeuGlyGlyGlyLysAspGlyIleGluProValSerProProGlnGlyMetTh	2232
Db	6606	GTCGGTCTTGGGTGGTGGTGAAGACGGTATTGAACCTGTGCCCCACCGAGGGCATAC	6666
Oy	2232	rGluProGlyHisSerArgSerAlaValTyProLeuLeuTyArgArgGlyGlnGlnTh	2252
Db	6666	GGAGCCAGGGGACTCCCGGAGTGTGTATACCGCTCGTGAACCGGAGTGGGAAACAGAC	6725
Oy	2252	rGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePh	2272
Db	6726	GGAGCCCAAGAGATGGAGCTCCAAAGTCTCCAGGGAACACCAAGCCAGCGCGCAAGCTTCTT	6785
Oy	2272	eSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGlnIleAsnLy	2292
Db	6786	CAGCAAGACTGACGAGAGCAACTCCGCGCATGTCTAAGTCCAAAGAACCAAGATCAACAA	6845
Oy	2292	sLysLeuAsnThrHisAsnArgAsnGluProGluTyTyrAsnIleSerGlnProGlyThrG	2312

Db 6846 GAAGCTGAACACCCACACCGGAAATGACCTGTAATACAAATATCAGCCAGCTGGAGCGA 6905
Qy 2312 uillePheanMeProAlaileThrglyleuMeThrTyArgSerGlnAlaVa 2332
Db 6906 GATCTTCAATATGCGCCGCATCACCGGAACAGGCTTATGACCTTATAGAAAGCAAGCGGT 6965
Qy 2332 lGngluHlsAlaSerThraMMeTglyleuGlnAlailelAArgLysAlaLeuMeTg 2352
Db 6966 GCAGGAACATGCGACGACCAACATGCGGCTGGAGGCAATATGAAAGGCACTATGG 7025
Qy 2352 ylyeTyraSPGlnTTPGluGlnSerProProleuSerAlaanaAlaPheanProleu 2372
Db 7026 TAAATATGACCAATGGGAAGATCCCGCGCTCAGCGCAATGCTTTAACTCTGAA 7085
Qy 2372 nAlaSerAlaSerLeuProAlaAlaMeCProlleThraAlaASPglYArgSerAspH 2392
Db 7086 TEGCACTGCGAGCTCCCGCTGCTATGCCCAATACCGCTGCTGAGAGAGAGTACCA 7145
Qy 2392 sThrleuThrSerProGlyGlyGlyGlyAlaLysValSerGlyArgProSerSerAr 2412
Db 7146 CACACTCACCTCGCGAGTGGCGCGGGAAGCCAAAGTCTCTGGCAGAGCCAGCAGCG 7205
Qy 2412 GlysaAlaLysSerProAlaProGlyleuAlaSerGlyAspArgProProSerAlaSerSe 2432
Db 7206 AAAAGCAAGTCCCGGCGCCCTGGCATGGAAGACGGACACCTCTGTCTCTC 7265
Qy 2432 rValHisSerGluGlyAspCyAsnaArgThrProleuThraAsnArgValTTPgluAs 2452
Db 7266 AGTGCACTGGAGGAGAGACTGCAACCGCGGAGCGGCTCAACACGGGTGGAGGA 7325
Qy 2452 pArgProSerSerAlaGlySerThraProPheProTyraAsnProleuileMeArgLeuG 2472
Db 7326 CAGGCGCTGCTCGCGAGTTCACAGCCATTCCTCAACCCCTGATCATGGCGTGA 7385
Qy 2472 nAlaGlyValMeAlaSerProProProProGlyleuProAlaGlySerGlyProleuAl 2492
Db 7386 GGGGGATGTCATGGCTTCCACCCCAACCGGCGCTCCCGCGGAGAGCGGCCCTCGC 7445
Qy 2492 aglyProHisAlaATrAspGluGluProLysProleuLeuCySerGlnTyrgluTh 2512
Db 7446 TGGCCCCCAACCAAGCTGGGAGAGAGAGCCAAAGCACTGCTGCTCGAGTACAGAGAC 7505
Qy 2512 rLeuSerAspSerGlu 2517
Db 7506 ACTCTCCGACAGCGAG 7521

RESULT 6
US-09-819-104A-1
Sequence 1, Application US/09819104A
Publication No. US20030027137A1
GENERAL INFORMATION:
APPLICANT: Chen, J. Don
TITLE OF INVENTION: NOVEL NUCLEAR RECEPTOR COMPRESSOR MOLECULES
TITLE OF INVENTION: AND USES THEREFOR
FILE REFERENCE: UMG-030
CURRENT APPLICATION NUMBER: US/09/819,104A
CURRENT FILING DATE: 2001-03-27
PRIOR APPLICATION NUMBER: 60/193,138
PRIOR FILING DATE: 2000-03-29
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 8686
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (157)..(7677)
US-09-819-104A-1

Alignment Scores:
Pred. No.: 0 Length: 8686
Score: 12978.00 Matches: 2487

Percent Similarity: 98.57% Conservative: 3
Best Local Similarity: 98.46% Mismatches: 9
Query Match: 98.21% Indels: 28
DB: 10 Gaps: 3
US-09-522-753-5 (1-2517) x US-09-819-104A-1 (1-8686)

Qy 1 MetSerGlySerThrglnLeuValAlaGlnThrTPArgAlaThGluProArgTyPro 20
Db 157 ATGTGCGGCTTCCACAGCCTTGCGACAGAGCTGAGAGGCCACATGAGCCGCTTACCCG 216
Qy 21 ProHisSerLeuSerThraProValGlnleAlaArgThraHisThraSPValGlyLeuLeu 40
Db 217 CCCACAGCCTTCTTACCCAGTGAAGATGCGCCGAGCACACGAGACGTCCGGCTCTCG 276
Qy 41 GluTyrglnHisHisSerArgAspTyraLaseHisLeuSerProGlySerileleGln 60
Db 277 GAGTACACAGCACACTCCCGGAGCTATGCTCCACCTGTCCGCCGCTCATCATCAG 336
Qy 61 ProGlnArgArgProSerLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
Db 337 CCCAGCGCGGAGGCGCTCCCTGCTGTGAGTTCAGCCCGGGAATGAACGCTCCAG 396
Qy 81 GluLeuHisLeuArgProGluSerHisSerTyraLeuProGluLeuGlyLysSerGluMet 100
Db 397 GAGCTTCACCTCGGCGAGATGCCACTATACCTGCCCCGAGCTGGGAAATCAGAGATG 456
Qy 101 GluPheileGlySerLysArgProArgLeuGluLeuLeuProAspProleuLeuArgPro 120
Db 457 GAGTTCATTGAAGAAAGAGCGCCCTCGGCTAAGAGCTGCTCTGACCCCTGCTGCGACCG 516
Qy 121 SerProleuLeuAlaThrglyGlnProAlaGlySerGluAspLeuThryAspArgSer 140
Db 517 TCACCCCTGCTGGCCACCGGCGAGCTGGGGATCTGAAGACCTCAACAAAGACCGTAC 576
Qy 141 LeuThrglyLysLeuGluProValSerProProSerProProHisThraSPProGluLeu 160
Db 577 CTGAGGAGCAAGCTGAACCGGATGTCCTCCCGAGCCCGCCGACACTACCTGAGCTG 636
Qy 161 GluLeuValProProArgLeuSerLysGluGluLeuileGlnAsnMeAspArgValAsp 180
Db 637 GAGCTGTGTCGCCCAAGGCTGTCAAGAGAGAGCTGATCAGAAACATGAGACCGCGTGAAC 696
Qy 181 ArgGluileThraMeValGluGlnGlnileSerLysLeuLysLysGlnGlnGlnLeu 200
Db 697 CGAGAGATCACCATGATGAGAGAGAGATCTTAAGCTGAAGAAAGAGAGCAACAGCTG 756
Qy 201 GluGluGlnAlaAlaLysProProGluProGluLysProValSerProProProleuGln 220
Db 757 GAGAGAGAGCTGCGCAAGCGCGCCGAGCTGAAGAGCCGATGACCGCCGCCCATCGAG 816
Qy 221 SerLysHisArgSerLeuValGlnleleTyraSPGluAsnArgLysLysAlaGluAla 240
Db 817 TCGAAGCACCGAGCTGTCAGATCATCTACAGAGAAACCGGAAGAAAGCTGAAGCT 876
Qy 241 AlaHisArgileLeuGluGlyLeuGlyProGlnValGluLeuProleuTyraSnglnPro 260
Db 877 GCACATCGGATCTGGAAGGCTGGGGGCCCGAGGTGAGCTCCGCTGTACAAACAGCCCC 936
Qy 261 SerAspThraArgGlnTyraHisGluAsnileLysileAsnGlnAlaMeArgLysLysLeu 280
Db 937 TCCGACACCCGCGAGATCATAGAAACATCAAAATTAACAGAGCGATGGGAAAGAGCTA 996
Qy 281 ileLeuTyraPheLysArgArgAsnHisAlaArgLysGlnTTPArgGlnLysPheCyGln 300
Db 997 ATCTTGTACTTCAAGAGAGGAATCACGCTCGGAACAATGGAGAGCAAGATTCTCCAG 1056
Qy 301 ArgTyraSPGlnLeuMeGlnAlaLeuGluLysLysValGluArgileGluAsnAspPro 320
Db 1057 CGCTATGACCACTCATGAGGCTGGGAGAAAGAGTGAAGCGCATTCAGAAACAACCCCC 1116
Qy 321 ArgArgArgAlaLysGluSerLysValArgGluTyraTyrgluLysGlnPheProGluile 340

Db 1117 CGCGCGCGGCCAAGAGAGAGAGTTCCGAGTACTACGAGAAAGCACTTCCCTGAGATC 1176
Qy 341 Arglysglnatrglyluwenglnl uargmecginserarValiglylnatrglysergly 360
Db 1177 CGCAAGCAGCCCGAGCTGCTCAGAGACCGCATGCG--AGGTTGGGCCAGCGGAGAGTGGG 1233
Qy 361 Leusermeserlaaalaargserglyuhisglnuvalserglulieiaaspglyleuser 380
Db 1234 CTGTCCATGTGCCCCCGCCGAGCGAGACGAGAGTGCAGAGATCATGATGCGCTCTCA 1293
Qy 381 GlnGlnGlnuamlenuglylvysglnmetArgglnleuvalaValileproPromeLeuTyr 400
Db 1294 GAGCAGAGAGAACTTGAGAAACAGATGCGCAGCTGGCCGTGATCCGCCCATCTGTAC 1353
Qy 401 AsplaaspGlnatrglylelyPheileamemcAnsglyleumelAaspProme 420
Db 1354 GAGCCTACCGAGCGCGCATTAATTCATCACTGAACGAGGCTTATAGCCGACCCCATG 1413
Qy 421 LysValTyrlysaasparGlnValmeAsmMetTpsersglnGlnlylvysglnThrPhe 440
Db 1414 AAGGTACAAAGACCCGACAGTGCATGACATGTGAGTGAAGAGAGAGAACCTTC 1473
Qy 441 ArgglylvysPheMetGlnhisProllyabnPhedglyleuileiaserPheleuglyarg 460
Db 1474 CGGAGAAAGTTCAAGCAGCATCCCAAGAACTTGGCTGATCGCATATCTCTGAGAGG 1533
Qy 461 LysThrValaIaglylvysValleuTyrTyrlyleuthrlylvysabnGlnuamTlyls 480
Db 1534 AAGACATGCTGATGCTGCTGCTCTATTACTACCTGACTGAAGAAATGAGAACTATAG 1593
Qy 481 SerleuValaIargArserTyrargArargglylvysersglnGlnGlnGlnGln 500
Db 1594 AGCTGTGAGACGAGACTATCGCGCGCGGCAAGAGCCAGCAACAGCAAGCAG 1653
Qy 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
Db 1654 CACAG 1713
Qy 521 AspGlylvysglnuvalserglnuvalaIaglylvysglnGlnlylvysProlGlnuam 540
Db 1714 GATGAG 1773
Qy 541 AspLyvglnuabPleuLeuLyvglnuvalserThrAspAspThrserGlyluabPasnAspGlu 560
Db 1774 GACAG 1833
Qy 561 LyvglnuvalaIaserylvysglnyarglylvysThrAlaanserglnGlnyarglylvys 580
Db 1834 AAGGAGGCTGTGCTCCAAAGGCGCAAACTGCCAAACAGCCAGGAGAGAGAGAGAG 1893
Qy 581 ArgIlethrArgserMetAlaansglnuvalaansserglnuvalaIleThrProlGln 600
Db 1894 CGCATACCCGCTCAATGGCTTAATGAGGCAACAGAGAGAGAGAGAGAGAGAGAG 1953
Qy 601 SerAlaGlyluuvalaaserMetGlnleuabnGlnuserSerArgTprThrGlnGlnMet 620
Db 1954 AGGCGGAGCTGGCTCTCATGAGAGCTGAATGAGGTTCTGCTGAGAGAGAGAGAG 2013
Qy 621 GlnThrAlaIalyslvysglnleuGlnuamhisglnyargAntTpseralileiaserMet 640
Db 2014 GAAAGAGCCCAAGAGAGTCTCTGAGAACAGCGCGCAACTGTGCGCATCGCCGATG 2073
Qy 641 ValGlySerlythrValaserGlnCyllysaabnPhelyrPheAntTlylvysPargln 660
Db 2074 GTGGGCTCCAGAGCTGTGTGAGTGAAGACTTCACTTCAACTACAGAGAGAGAG 2133
Qy 661 AsnleuabPglulileuGlnGlnhislvyslvyslvyslvyslvyslvyslvyslvys 680
Db 2134 AACCTCGATGAGATCTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2193
Qy 681 Arglylvyslvyslvyslvyslvyslvyslvyslvyslvyslvyslvyslvyslvys 700
Db 2194 AGGAG 2253

Qy 701 GlnabPgluGlnMetGlnAlaSerGlyValserGlylvysenglnGlnMetValGlnGln 720
Db 2254 GAGATAGAGAGATGAGAGGCTCGGCTGACGAGAGAGAGAGAGAGAGATGAGAGAG 2313
Qy 721 AlaGlnAlaLeuHisAlaserGlylvysGlnuvalProlargGlylvysserGlyProla 740
Db 2314 GCTGAA-----GCC 2332
Qy 741 ThrValAsnanserSerAspThrGlnSerileProserProHisThrGlnAlaAllys 760
Db 2323 ACTGTCAACAACTGACCTAGACACGAGAGATCCCTCTCTCTCACATGAGCGCCAG 2382
Qy 761 AspThrGlylvysenglylvysProlAspProlaIleThrleuglylvysabPglProlPro 780
Db 2383 GACACAGGAGAGATGGAGCCCAAGCCCAAGCCCACTTGGAGCCCAAGAGAGAGAG 2442
Qy 781 GlyProlProThrProPro-ArgArgThrSerArgAlaProlleGlnProlProAlase 800
Db 2443 GGGCCACCAACCCAGCAGCAGAGAGATCCCGCC-CCCACTGATCCACCCCGGCTC 2501
Qy 800 rGlnuAlaThrGlyAlaProlThrProProlaIaProlProSerProSerAlaProlPro 820
Db 2502 TGAAGCCACTTGAAGCTTACGCCCCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2561
Qy 820 ovalValProlleuglnuvalserglnuvaluThrAlaAlaProlProValGlnGln 840
Db 2562 TGTGTCTCCAG 2621
Qy 840 YGlnGlnGlnlylvysProlaIaIaglnGlnleuAlaValAspThrGlylvysAlaGln 860
Db 2622 GAG 2681
Qy 860 uproValysSerGlylvysThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 880
Db 2682 GCCGTCAAG 2741
Qy 880 aglnuAlaIaglnuAlaThrAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 900
Db 2742 GAGAGCGCTGAG 2801
Qy 900 rGlyArgAlaThrThrAlaIalysSerSerGlyAlaProlGlnAspSerAspSerAlaThr 920
Db 2802 CGGAG 2861
Qy 920 rCyserAlaAspGlnuvalaIaglnuvalaIaglnuvalaIaglnuvalaIaglnuval 940
Db 2862 CTCAGGTGAG 2921
Qy 940 oArgProlSerleuLeuThrProlThrGlyAspProlaIaAsnAlaserProlGlnysPr 960
Db 2922 AAGGCCAGCTCTCTCAACCCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 2981
Qy 960 oleuAspLeuLyvglnuvalserGlnuvalaIaIaIaIaIaIaIaIaIaIaIaIa 980
Db 2982 ACTGGAAGCTGAG 3041
Qy 980 sValHisGlnupProProlargGlnuabPalaIaProlThrlylvysProlaIaProl 1000
Db 3042 AGTCATGAG 3101
Qy 1000 oProlProGlnAsnleuGlnProlGlnserAspAlaProlGlnGlnProlGlnysSer 1020
Db 3102 GCCACCGCAAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3161
Qy 1020 gGlylvysSerArgSerProlaProlProlaIaAspLyvgln----- 1033
Db 3162 GGGCAAG 3221
Qy 1034 -----AlaPheAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 1052
Db 3222 CCCAGCTTGTGAG 3281

1052 yLeuProPheProValProProArgGluValAlaIleValAspProHisAlaProAspPr 1072
|||||
3282 CCTGCCCTTCCCGTCCCGCGTAGAGGATCAAGGCGCTCCCGCATGCCCCGAGACC 3341
|||
1072 oSerAlaPheSerTyrAlaProProGluYhiSPProLeuProLeuGlyLeuHisAspThrAl 1092
|||||
3342 CTCAGCCTTCTCTTAAGCTCCACCTGTATCAACCACGTGCGCTCGGCTCATAGACTGC 3401
|||
1092 aArgProValLeuProArgProProThrIleSerAspProProProLeuIleSerSerAl 1112
|||||
3402 CCGGCCCCGCTCTCCCGCGCCACCCACATCTCCAACCCGCTCCCTCATCTCTCTGC 3461
|||
1112 alyshisProSerValLeuGluValGlnIleGlyAlaIleSerGlnGlyMetSerValG 1132
3462 CAAGCACCCACAGGCTCTCGAGAGGCAAAATAGGTGCATCTCCCAAGAAATGCGGTCCA 3521
|||
1132 nLeuHisValProTyrSerGluHisAlaIlyValaProValGlyProValThiMetGlyLe 1152
|||||
3522 GCTCCACGTCCTCTACTCAGAGCATGCCAAGCGCCGGTGGGCTGTCAACATGGAGGCT 3581
|||
1152 uProLeuProMetAspProLyshLeuAlaProPheSerGlyValIlyGlnGlnLe 1172
3582 GCGCCCTGCCATAGACCCCAAAAGCTGGCACCTTTCAGGGAGTGAAGCAGAGCATCT 3641
|||
1172 uSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAl 1192
3642 GTCCCCACGGGGCCAGGCTGGGCCACCGGAGAGCTGGGGGTGCCACAGCCAGAGGCG 3701
|||
1192 aSerValLeuArgGlyYThrAlaLeuGlySerValProGlyGlySerIleThiArgGlyI 1212
3702 GTCCGGTCTGAGAGGACAGACTGTGGGCTCATGTTCCGGGGGGAAGCATCCAAAGCAT 3761
|||
1212 eProSerThrArgValProSerAspSerAlaIleThrYrArgGlySerIleThiArg 1232
3762 TCCACGACACCGGGTCCCTCGACAGCGGCATCACTACCGGGCTCCATCCACCG 3821
|||
1232 yThrProAlaAspValLeuTyrIlyGlyYThrIleThiArgIleIleGlyGluAspSerPr 1252
3822 CAGCGCAGCTGAGCTCTGTATCAAGGCAACATCAACAGGATCATCGCGAGAGACAGCCC 3881
|||
1252 oSerAlaGluAspArgGlyYArgGlyuAspSerLeuProLyshIlyHisValIleTyrGlu 1272
3882 GAATCCTCTTGAGCGCGGCGGAGGACAGCCTGCCCAAGGCGCATCTCTTCAAGAA 3941
|||
1272 yLyshLyshGlyHisValLeuSerTyrGlnGlyGlyMetSerValThiGlnCysSerIlyArg 1292
3942 CAAGAAAGGCCAAGCTCTTGTCTATAGAGGTGGATGTCTGTACCCAGAGTCCAAAGA 4001
|||
1292 uAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLyshArgThrTyrAs 1312
4002 GGAAGGACAGAGCAGCTCAAGACCCCCCATGAGAGCGCGCCCAAGCGCACCTATGA 4061
|||
1312 pMetMetGluGlyYArgValGlyYrArgAlaIleSerSerAlaSerIleGlnGlyLeuMetG 1332
4062 CAGGATGAGAGGCGCGGTGGGAGAGCATCTCTCAAGCAGATCAAGAGTCTCAAGGG 4121
|||
1332 yArgAlaIleProProGluYrArgHisSerProHisHisLeuLyshGlnGlnHisIleArg 1352
4122 CCGTCCCATCTCCCGGAGGACACAGCCCCCACACCTCAAAAGACAGACACATCCG 4181
|||
1352 gGlySerIleThiGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArg 1372
4182 CGGCTCATACACAAGAGGATCCCTCGGTCTAGTGAGGACAGAGAGACTACCTGCG 4241
|||
1372 gArgGluAlaIlyLeuLeuLyshArgGlyGlyYThrProProProProProSerArgAs 1392
4242 TCGGAGGCGCAAGCTCTAAAGCGGAGGACAGCGCTCCGCCCCACCGGCTTCACGGA 4301
|||
1392 pLeuThrGluAlaTyrIlyThiGlnAlaLeuGlyProLeuLyshLeuYrProAlaHisG 1412
4302 CTTGACGAGGCTTCAAGACGAGCGGCTGGGCGCCCTTAAGCTGAAGCGGCGCATGA 4361
|||
1412 uGlyLeuValAlaThrValLyshGluAlaGlyArgSerIleHisGluIleProArgGlu 1432

4362 GGGCCTGTGGCCACCGTGAAGAGCGGGCGCTCATCCATGAGATCCCGGACAGA 4421
|||||
1432 uLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLyshGlySerIleTh 1452
4422 GCTGGGACACAGCGCGGAGCTGCCCTTGAGCCCGCGCGGCTCAAGAGGCTCATAC 4481
|||
1452 rGlnGlyYThrProLeuLyshAspThrGlyAlaSerThiThiGlySerIlyshHisAs 1472
4482 GCAGGGCACCCGCTCAAGTACGACACCGCGCGTCCACCATGGCTCCAAAAAGCACGA 4541
|||
1472 pValArgSerLeuIleGlySerProGlyYrArgHisPheProProValHisProLeuAspVa 1492
4542 CGTACGCTCTCTATCGAGGCCCGCGGAGGTTCCACCGGTGACCCGCTGAGATGT 4601
|||
1492 lMetAlaAspAlaArgAlaLeuGluYrArgAlaCysTyrGlnGluSerLeuLyshSerArgPr 1512
4602 GATGCCGACGCGCGGACCTGGAAAGCTGCTCTCTCAAGAGAGCTGAAGAGCGGGCC 4661
|||
1512 oGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProG 1532
4662 AGGAGCGGCACAGACTCGGGGGGCTCCATTGGCGCGCGCGCCGCTCATGTGCTTG 4721
|||
1532 uLeuGlyYrProArgGlnSerProLeuThiYrArgIlyAspHisGlyAlaProPheAlaG 1552
4722 GCTGGGTAAAGCGCGGACAGGCCCTCTGACCTATGAGAACAGGGGACACCTTGCGG 4781
|||
1552 yHisLeuProArgGlyYrProValThiMetArgGluProThrProArgLeuGlnGlu 1572
4782 CCACTCCCAACAGATCTCCCGTATCAATGGGGAGGCCACGCGGCTGAGAGAGG 4841
|||
1572 ySerLeuSerSerSerLeuAlaSerGlnAspArgIlyshLeuThiSerThiProArgGlu 1592
4842 CAGCTTTGTGTCAGAGAAAGCATCCAGGACGGAAGCTGACAGCTCCATGATAGAT 4901
|||
1592 eAlaIlySerProHisSerThiValProGluHisHisPheHisProIleSerProTyrG 1612
4902 CGCCAGTCCCGGACAGGACCGTGCAGAGCACCCACACCCATCTCGCTTATGA 4961
|||
1612 uHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPh 1632
4962 GCACTGTCTCGGGGAGTGAAGTGGCTGAGACCTGTATCGAGCAGATCCCTTGCGCTT 5021
|||
1632 eAspProThiSerIleProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuPr 1652
5022 CAGCCCACTCATACCCCGGAGCATCCCTGTGAGGAGCGCGCTTACTTACTGCC 5081
|||
1652 oArgHisLeuAlaProAspProThiYrProHisLeuTyrProProTyrLeuIleArgG 1672
5082 CCGACACTGGCCCCCAACCCCACTTACCGGACCTGTACCCACTTACTTACCGGG 5141
|||
1672 yTyrProAspThrAlaAlaLeuGluYrAspArgGlnThiIleAspAspTyrIleThiSe 1692
5142 CTACCCCGACAGCGCGCGCTGAGAAACCGGACGACATCATCAATGACTACATCACTC 5201
|||
1692 rGlnGlnMetHisHisAspThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgG 1712
5202 GCAGCAGATGACCAACAACGGGCAACCCCAATAGGCCGAGCGAGTATATGCTGAGGG 5261
|||
1712 yLeuSerProArgGluSerSerLeuAlaLeuMetTyrAlaAlaGlyProArgGlyIleI 1732
5262 CCTCTGCCCCCGGAGATCTCGTGGCACTCACTAGCTGCGGGTCCCCGAGGAGCAT 5321
|||
1732 eAspLeuSerGlnValProHisLeuProValLeuValaProProThiProGlyThrProAl 1752
5322 CAGCTGTCCCAAGTGCACACTGCTGTGTGTGTGCCCCGACACAGAGACCCAGC 5381
|||
1752 aThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHis 1772
5382 CACCGCATGAGACGCTTGCTTACTCCCAACCGGCGCCCAAGCCCTTACGAGCGCCCA 5441
|||
1772 sSerSerSerProLeuSerProGlyGlyProThiHisLeuThiYrProThiThiThiSe 1792

Db 5442 CACGAGCTCCCACTCTCCCGAGAGGTCCAAACACTTGACAAACCAACACACGCTC 5501
Qy 1792 rSerSerGluArgGluArgAspArgAspArgGluArgAspArgGluArgGluArg 1812
Db 5502 CTGGTCGAGACGGAGGAGACCGGGATCCAGAGCGGAGCCGGATCGGAGCGGAGAAA 5561
Qy 1812 sSerIleuThrSerThrThrValGluHisAlaProIleTPrArgProGlyThrGly 1832
Db 5562 GTCCATCTCAGCGTCCACCAAGAGGTGAGACGACCACTCTGAGACTGTGATCAGA 5621
Qy 1832 uGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgProAl 1852
Db 5622 GCAGAGAGCGGAGCGAGCGAGCGAGCGGGGGTGGGGGCGAGCGCGCGCGCGC 5681
Qy 1852 aSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspIleAlaGlu 1872
Db 5682 CTCCCACTCCCAATGCCACCAAGCACTGCCCATCTCCCTCGAGACCAAGATGCCCTCCA 5741
Qy 1872 nglnArgProSerValIleuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluPr 1892
Db 5742 GCAGAGACCCAGTGTCTTCAACAACAGGCAATGAGGTATCATCACCGCTGTGAGGCC 5801
Qy 1892 oSerLysProThrValIleuArgSerThrSerThrSerSerProValArgProAlaAlaThr 1912
Db 5802 CACCAAGCCACAGGTCTGTGAGGTCCACCTCCACCTCTCAACCGTTGCCAGCTGCAC 5861
Qy 1912 rPheProProAlaThrHisAcysPProLeuGlyGlyThrLeuAspGlyValTyrProThrLe 1932
Db 5862 ATTCACCACTGCCACCACTGCCACTGGGGGCACTCGATGGGGTCTAACCTTACCT 5921
Qy 1932 uMetGluProValIleuLeuProLysGlyAlaProArgValAlaArgProGluArgProAr 1952
Db 5922 CATGAGAGCCGCTGTGCTGCCAAGAGGCCCCCGGGTCCCGGCGCAAGGGGGCCCG 5981
Qy 1952 gAlaAspThrGlyHisAlaPheIleuAlaLysProProAlaArgSerGlyLeuGluProAl 1972
Db 5982 AGAGAGACACGGGCAATGCTTCTCGCAAGCCCGCGCTCCGGCTGAGAGCGCGC 6041
Qy 1972 aSerSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAl 1992
Db 6042 CTCTCTCCCAAGAGGGCTCGAGGCCCGGCCCTTAGTCTCTCTCTCTGACCAAGC 6101
Qy 1992 aThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProP 2012
Db 6102 CACCACTGCGCGCACCCCTCGAAGAACTCGCACTCACACGCGCAAGCCGAGCCGCC 6161
Qy 2012 oAlaProProAlaSerAlaSerAspProHisAlaArgGluLysThrGlnSerLysPhePhe 2032
Db 6162 GGGCGCACTGCTCGGCTCGGACCGGACCGGAGAAAGACTCAAGTAAACCTTTTTC 6221
Qy 2032 rIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluG 2052
Db 6222 CATCCAGAACTGGAACTCCGTTCTTGAGTTACACGCGAGAGACTTACAGCCCGAAG 6281
Qy 2052 yValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLys 2072
Db 6282 GGTGGAGCCCGTCAGCCCTGTGACTCACCACTGACCAAGCAAGAGGCTCCCA 6341
Qy 2072 sHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProG 2092
Db 6342 GAACCTGGAAGAGCTGCAAGAGCACTGAGGGGGAGCTCGGCCCAAGAGCGCAGG 6401
Qy 2092 yProValLysLeuGlyGlyValAlaAlaHisLeuProHisLeuArgProLeuProGluS 2112
Db 6402 CCCCCTGAAGCTTGGGGGGAGGCGCCCACTCCCAACACTGGCGGCTGCTAGAG 6461
Qy 2112 rGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgVa 2132
Db 6462 CCAGCCCTGCTGTCAGGCGGCTGCTCCAGACCGCCCAAGGGGTCAAGGTCAACAGGG 6521
Qy 2132 lValThrLeuAlaGlnHisAlaSerGluValIleThrGlnAspTyrThrArgHisHisP 2152
Db 6522 GGTCACTTGGCCGACACATCATGAGGTATCATCAAGAGCTTACACCGGAGCACACCC 6581

Qy 2152 oGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysPr 2172
Db 6582 ACAGAGCTACAGCCACCCCTGCCCCGCCCTCTACTCTCTCTCTCTCTCTCTCTCTCT 6641
Qy 2172 oValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHisGlyAl 2192
Db 6642 CGTCTTGAACTCCGCGCGCCCAACAGTGAACCTTCACTCCCGCGCGGAGCATGTGTGC 6701
Qy 2192 aProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnLysThr 2212
Db 6702 CCGGCGCGGTGGCTCCCCCAAGCGAAGGGGGCAAGAGGTCTTCAAGCCAAACAAGAC 6761
Qy 2212 rSerValLeuGlyGlyGlyLysArgGlyIleGluProValSerProProGluGlyMetThr 2232
Db 6762 GTCGGTCTTGGGTGTGTGTAGAGCGTATTGAACTGTGTCTTCCCAACGAGGGCATGAC 6821
Qy 2232 rGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnTh 2252
Db 6822 GAGAGCGAGGACCTCCGAGAGTGTGTGATCCGCTGTATCCGGGATGGGAAACAGAC 6881
Qy 2252 rGluProSerArgMetGlySerLysSerProGlyLysAsnThrSerGlnProProAlaPheP 2272
Db 6882 GAGGCCAGAGGATGGGTCTCAAGTCTCCAGGCAACACAGCGCAGCGCAGCTTCTT 6941
Qy 2272 eSerLysLeuThrGlySerSerSerAlaMetValLysSerLysLysGlnIleAsnLys 2292
Db 6942 CACCAAGTACCGAGAGCACTCCGCAATGTGTCAAGTCCAAAGAGCAAGATCAACAA 7001
Qy 2292 sLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrG 2312
Db 7002 GAAGCTGAACACCAACACCGAATGAGCTGATTAATCAATATCAGCAGCTGGAGCGA 7061
Qy 2312 uIlePheAsnMetProAlaIleThrGlyLysMetThrTyrArgSerGlnAlaVa 2332
Db 7062 GATCTTCAATATGCCCCGCAATCAACGAAACAGGCTTATGACTTAAGAACCAAGGGGT 7121
Qy 2332 lGlnGlnHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetG 2352
Db 7122 GCAGGAACATCCAGACCAACCAATGGGGCTGGAGGCAATTAATGAAGGCACTCATGG 7181
Qy 2352 yLysTyrAspGlnTyrGluGluLysSerProLeuSerAlaAsnAlaPheAsnProLeuAs 2372
Db 7182 TAAATATGACAGATGGGAAGAGTCCCCGCGCTCAAGGCCAATGCTTTTAACTCTGAA 7241
Qy 2372 nAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHis 2392
Db 7242 TGCAGTGCAGGCTTCCGCTGTATGCCATMAACCGTGTGAGAGGAGGTACCA 7301
Qy 2392 sThrLeuThrSerProGlyGlyGlyLysValLysValSerGlyArgProSerSerAr 2412
Db 7302 CACACTCAGCTGCCAGGTGGCGGGAGGAGGCAAGGTCTCTGGAGACCCAGACCG 7361
Qy 2412 gLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSe 2432
Db 7362 AAAAGCAAGTCCCCGCGCCCGGCTGTGATCTGGGAGCCGCGCAACCTCTCTCTC 7421
Qy 2432 rValHisSerGluGlyLysPcyAsnArgArgThrProLeuThrAsnArgValTyrGluAs 2452
Db 7422 AGTGCACTGGAGGGAGACTGCAACCGCGAGACGCGCTCACCAACCGGCTGTGGAGGA 7481
Qy 2452 pArgProSerSerArgLysThrProPheProTyrAsnProLeuIleMetArgLeuG 2472
Db 7482 CAGGCGCTGTGCTCGAGGTTCACAGCAATTCCTTCAACCCCTTATCATGCGGCTGCA 7541
Qy 2472 nAlaGlyValMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeuAl 2492
Db 7542 GGGGGGTGTCTATGGCTTTCCCAACCCCAACGGAGCTCCCGGGGAGGAGCGGCGCTCG 7601
Qy 2492 aGlyProHisHisAlaTPrAspGluGluProLysProLeuLeuCysSerGlnTyrGluL 2512
Db 7602 TGGCCCCCAACACGCTGGGAGAGAGAGCCAAAGCCATGCTGTGCTGCAGTACGAGAC 7661

QY 2512 rleuSerapSerGlu 2517
DB 7662 ACTCTCGACAGCGAG 7677

RESULT 7

US-10-174-014-4
Sequence 4, Application US/10174014
Publication No. US20040005292A1
GENERAL INFORMATION:
APPLICANT: C. Frank Bennett
APPLICANT: Susan M. Freiler
APPLICANT: Kenneth W. Doble
TITLE OF INVENTION: ANTISENSE MODULATION OF SMRT EXPRESSION
FILE REFERENCE: PTS-0012
CURRENT FILING DATE: 2002-06-17
NUMBER OF SEQ ID NOS: 73
SEQ ID NO 4
LENGTH: 8686
TYPE: DNA
ORGANISM: H. sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (157)...(7680)
US-10-174-014-4

Alignment Scores:

Pred. No.: 0 Length: 8686
Score: 12978.00 Matches: 2487
Percent Similarity: 98.57% Conservative: 3
Best Local Similarity: 98.46% Mismatches: 9
Query Match: 98.21% Indels: 28
DB: 18 Gaps: 3

US-09-522-753-5 (1-2517) x US-10-174-014-4 (1-8686)

QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTPARAlaThrGluProArgTyrPro 20
DB 157 AGTGGGGCTCCACACAGCCTGTGGACACAGTGGAGGGCCACTGAGCCCGCTACCGG 216
QY 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
DB 217 CCCACAGGCTTTCCTACCAAGTCAGATCGCCGGACCAACGAGACGTCGGCTCTG 276
QY 41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
DB 277 GAGTACACACACACACTCCCGCAGCTATGCTCCCACTGTCGCCGCTCCATCATCCAG 336
QY 61 ProGlnArgArgProSerLeuSerGluPheGlnProGlyValAsnGluArgSerGln 80
DB 337 CCCACAGGCGGAGGCGCTCTGCTGTGAGTTCAGCCCGGAAATGACGATCCAG 396
QY 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLeuSerGluMet 100
DB 397 GAGCTCCACTCGGCGCAGAGTCCCACTATACCTCCGAGCTGGGAAAGTCAGAGATG 456
QY 101 GluPheIleGluSerIleArgProArgLeuGlnLeuLeuProAspProLeuLeuArgPro 120
DB 457 GAGTTCATTGAAGACGAGCGCTCGCTAGAGACTCTGCTGACCCCTGCTGCACCG 516
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrIleAspArgSer 140
DB 517 TCACTCCCTGTGGCCACGGGCGAGCTGCGGATCTGAAGACTCAACCAAGACCGTAGC 576
QY 141 LeuThrGlyLeuLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
DB 577 CTGACGGGCAAGTGGAAACGGGTGTCTCCCGCAGGCCCCCGCACACTGACCTGAGCTG 636
QY 161 GluLeuValProProArgLeuSerIleGluGlnLeuIleGlnAsnMetAspArgValAsp 180
DB 637 GAGCTGTGCGCCGACAGCGGTGTCCAAAGAGAGACTGATCCAGAAACATGACCGCGTGAC 696

QY 181 ArgGluIleThrMetValGluGlnGlnIleSerIleLeuValIleGlyGlnGlnLeu 200
DB 697 CGAGGATCACCATGTGTAGACACAGATCTTAAGCTGAAGAAAGACACAGACGCTG 756
QY 201 GluGluGluValAlaIleValProProGluProGluIleValSerProProIleGln 220
DB 757 GAGGAGGAGGCTGCACAGCCGCCGAGCTGAGAAAGCCGGTGTCCGCCGCCATTCAG 816
QY 221 SerIleHisArgSerLeuValGlnIleIleTyrAspGluAsnArgIleValAlaGlnIle 240
DB 817 TCGAAGCACCCGACGCTGTGCAATCATCTACAGACGAAACCGGAAAGGCTGAGCT 876
QY 241 AlaHisArgIleLeuGlnGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPro 260
DB 877 GCACATCGGATTTGGAAGGCTCGGGGCCAGGGGAGGCTGCCCTGTACCAACAGGCC 936
QY 261 SerAspThrArgGlnTyrHisGluAsnIleIleValIleAsnGlnAlaMetArgIleValLeu 280
DB 937 TCCGACACCCGCACTATCATGAAACATCAAAATTAACACAGGCATCGGAAAGACTA 996
QY 281 IleLeuTyrPheIleValArgArgAsnHisAlaArgIleGlnTyrGlyLeuValIleValPheCysGln 300
DB 997 ATCTTTGTACTTCAGAGAGAGATACCGCTCGAAACATGGAGACAGAGTTCTGCCAG 1056
QY 301 ArgTyrAspGlnLeuMetGluValAlaLeuGluIleValIleValIleGluValIleGluValAsnPro 320
DB 1057 CGCTATGACACAGCTCATGAGAGCGCTGGGAGAAAGGTGGAGCGCATCGAAGAACACCC 1116
QY 321 ArgArgArgAlaValGlySerIleValIleArgGluTyrTyrGluValGlnPheProGluIle 340
DB 1117 CGGCGCGCGGCGAAGAGAGCAAGGTGCGAGTACTACGAGACAGATTCCTTGAGATC 1176
QY 341 ArgIleValIleValGlnLeuGlnIleValArgMetGlnSerArgValIleGlnArgIleGlySerGly 360
DB 1177 CGCAAGACGCGGAGCTCAGGAGAGCGCATGCG--AGGTTGCGCAGCGGGGCGAGTGG 1233
QY 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspIleLeuSer 380
DB 1234 CTGTCCATGTCCGCCCGCGCCGCGCAGACAGAGGTGTCAAGATATCATGAGCTCTCA 1293
QY 381 GluGlnGlnIleValLeuGlnIleValArgGlnLeuAlaValIleProProMetLeuTyr 400
DB 1294 GAGCAGAGAACTTGAGAGAGAGATGCGCAGCTGGCCGTATCCGCCCATGCTGTAC 1353
QY 401 AspAlaAspGlnGlnIleArgIleValSerPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420
DB 1354 GACGCTGACACAGACCGCATCAATCAATCAACAGAAACGGCTTATGGCCGACCCCATG 1413
QY 421 IlyValIlyIleValAspArgGlnValIleValMetTyrPheGlnGlnIleValIleValThrPhe 440
DB 1414 AAGGTGTACAAAGACCGCCAGGTCATGAACATGTGAGTGAACAGAGAAAGAGACCTTC 1473
QY 441 ArgGluValPheMetGlnHisProIleValAsnPheGlyLeuIleAlaSerPheLeuArg 460
DB 1474 CGGAGAAAGTTATCATGAGATCCCAAGAACTTTGGCTGTATGTCATCTCTGAGAGAG 1533
QY 461 IlyThrValAlaGlyCysValIleuTyrTyrTyrLeuThrIleValAsnGluValAsnTyrIle 480
DB 1534 AAGACAGTGTGAGTGTGCTCTCTATTACTGACTCAAGAAAGATGAGAACTATAAG 1593
QY 481 SerLeuValIleArgArgSerTyrArgArgArgIlyIleValSerGlnGlnGlnGlnGln 500
DB 1594 AGCTGTGTAGACGAGGCTATGCGGCCCGCCGCAAGACCAACAGAAACGACGACG 1653
QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
DB 1654 CAGCAGACAGACAG 1713
QY 521 AspGluIleValIleValIleValIleValIleValIleValIleValIleValIleValIleVal 540
DB 1714 GATGAG 1773
QY 541 AspIlyGluAspLeuLeuIleValIleValIleValIleValIleValIleValIleValIleVal 560

D	1774	GACACGAGAAAGCCTTCCTCAAGAGAAAGACACACCTCAGGGGAGGACAAAGACGAG- 1833	D	2802	CGGCGAGGCCACCAACGACCAAGACTCGGGCGCCCCCAGGAGACGACCTCCAGTCCAC 2861
Q	561	LYSGIUAIAVALAIASerLYSGIARGLYThrAlaAsnSerGlnGlyARGARGLY 580	Q	920	rcYSerAlaAspGluValAspGluAlaGluGlyIAspLYSAsnArgLeuLeuSerPr 940
D	1834	AAGAGAGCTGTGGCTCCAAAGGCGGCAAACTCCAAACAGCCGAGGAAGACGCAAGGCG 1893	D	2862	CTGCAGTGCAGACAGAGGTGAGTGGCGAGGGGGGAGCAAGAAACGGCTGTCTCCC 2921
Q	581	ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600	Q	940	oArgProSerLeuLeuThrProThrGlyYAspProArgAlaAsnAlaSerProGlnYsPr 960
D	1894	CGCATCACCCGGCTCAATGGCTAATGAGGCCAAACAGCGAGGCGCATCATCCCCACAG 1953	D	2922	AAGGCCAGCTCCTCAACCCGACTGGCGACCCCGGGGCCAATGCTCAACCCAGAAAGCC 2981
Q	601	SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluMet 620	Q	960	oLeuAspLeuLYSGIleuLYSGIleuArgAlaAlaAlaIleProProIleGluValThrLY 980
D	1954	AGCCCCAGCTGGGCTCCATGAGACTGAATGAGATTCTCGCTGAGACGAAGAAATG 2013	D	2982	ACTGACCTCAAGACGCTGAAGACGAGCGGCTGGCCATCCCCCATCCAGGTCAACAA 3041
Q	621	GluThrAlaLYSlySGIleuLeuGluIleGlyArgAsnTrpSerAlaIleAlaArgMet 640	Q	980	sValIleGluProProArgGluAspAlaAlaProThrLYSProAlaProProAlaProPr 1000
D	2014	GAAACACGCCAAGAAAGGTCTCTGAAACAGCGCGCAACTGGTGGCCATCGCCGAGATG 2073	D	3042	AGTCCATGAGCCCCCGGGAGAGACGAGCTCCCAACCAACAGCTCCCCAGGCCAC 3101
Q	641	ValGlySerLYSThrValSerGlnCysLYSAsnPhetYrPheAsnTrpLYSlyARGln 660	Q	1000	oProProGluAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProAr 1020
D	2074	GTTGGGCTCCAAAGACTGTGTGGCACTGTAAAGACTTCTAATTCAACTCAAGAAAGGCG 2133	D	3102	GCCACCGCAAAACCTGCAGCCGAGAGCCAGCCCTCCACAGCTGGCAGACGCCCG 3161
Q	661	AsnLeuAspGluIleLeuGlnGlnIleLYSleuLYSmetGluLYSGluArgAsnAlaArg 680	Q	1020	gGlyLYSserArgSerProAlaProProAlaAspLYSGlu----- 1033
D	2134	AACCTCATGATGATCTTGACAGACACAGCTGAAGATGAGAGAGAGAGACGCCGCG 2193	D	3162	GGGCAAGACGACGAGCCCGGACCCCGCCGACCAAGAGAGACAGAACCTGTCTT 3221
Q	681	ArgLYSlySerLYSlySAlaProAlaAlaAlaSerGluGluAlaAlaPheProProVal 700	Q	1034	-----AlaPheAlaAlaGluValaGluIleuLYSleuProGlyYAspProProCysTrpThrSerG 1052
D	2194	AGGAAAGAAAGAAAGGCGCGCGCGCGCGCAGCGAGAGGTGATTCGCCGCCGTGGTG 2253	D	3222	CCAGGCTTGACAGCGGAGCCCAAGAGCTGCTGGGAGACCCCTTGCTGTGACTTCGG 3281
Q	701	GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGlu 720	Q	1052	LYSleuProPheProValProProArgGluValIleLYSAlaSerProIleAlaProAspPr 1072
D	2254	GAGGATAGAGAGATGAGAGGCGCTCGGGCTGACGGGAATGAGAGAGATGTGTGAGAG 2313	D	3282	CTGCGCTTCCCGCTGCCCGCCCGGTAGAGTATCAAGGCTTCCCGCATGCCCGGACCC 3341
Q	721	AlaGluAlaLeuIleAlaSerGlyAsnGluValProArgGlyGlyCysSerGlyProAla 740	Q	1072	oSerAlaPheSerTYrAlaProProGlyIleIleProLeuProLeuGlyLeuIleAspThrAl 1092
D	2314	GCTGAA-----GCC 2322	D	3342	CTCAGGCTTCTCTCAAGCTTCACCTGATCCACCTGAGCCCTCGGGCTCCATACACTGC 3401
Q	741	ThrValAsnAsnSerSerAspThrGluSerIleProSerProIleThrGluAlaAlaLYS 760	Q	1092	aArgProValLeuProArgProProThrIleSerAsnProProLeuIleSerSerAl 1112
D	2323	ACTGTCAACAACAGCTAGACACCGAGAGATCCCTCTCTCTCACTGAGGCGCGCAAG 2382	D	3402	CGGCGCGTCTCGCGGCCACCAACACTTCAACCCGCTCCCTCATCTCTCTG 3461
Q	761	AspThrGlyGluAsnGlyProLYSProProAlaThrLeuGlyAlaAspGlyProProPro 780	Q	1112	alYsHISProSerValLeuGluIleArgIleGlyAlaIleSerGlnLYSmetSerValG 1132
D	2383	GACACAGGGGAGAAATGGGCCCAAGCCCCCAGCCACTTGGGCGCGGACGCCCA 2442	D	3462	CAGACACCCAGGCTCTCGAGAGGCAAAATAGTGCATCTCCCAAGAAATGTCTGCTCA 3521
Q	781	GlyProProThrProPro-ArgArgThrSerArgAlaProIleGluProThrProAlaSe 800	Q	1132	nLeuHISvalProTYrSerGluIleAlaLYSAlaProValGlyProValThrMetGlyLe 1152
D	2443	GGGCGACCCACCCACCAACCGAGAGACATCCCGGCC-CCCACTGAGTCCACCCCGGCTC 2501	D	3522	GCTCCACGTCTCCGTACTCAGAGCATGCCAAAGGCCCGGTGGGCTGTGTACCATGGGGCT 3581
Q	800	rgIuAlaThrGlyAlaProThrProProProProAlaProProSerProSerAlaProProPr 820	Q	1152	uProLeuProMetAspProLYSlySleuAlaProPheSerGlyValIleGluGlnGluIle 1172
D	2502	TGAAGCACCTTAGCCCCCTAGCGCCCAACGACCCCACTTCCCTTCACTCTCTCC 2561	D	3582	GCCCTGCGCCATGAGACCCCAAAAGCTGGACCTTCAAGCGAGTAAAGAGAGACGCT 3641
Q	820	oValValProLYSGluGluLYSGluGluGluIleThrAlaAlaAlaProProValGluGlu 840	Q	1172	uSerProArgGlyGluAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAl 1192
D	2562	TGTGTCTCCCAAGAGAGAGAGAGAGAGACCGCAGCGAGGCCCACTGTGAGAGAGGG 2621	D	3642	GTCGCCACGGGGCAGAGCTGGGCCACCGAGAGCTGGGGGTGCCACAGCCCAAGAGAGC 3701
Q	840	YGIuGluGluLYSProProAlaAlaGluGluLeuAlaValAspThrGlyLYSAlaGluG 860	Q	1192	aSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLYSlyI 1212
D	2622	GGAGGAGCAAAACCCCCCGGCTGAGAGCTGGCATGTGACACAGCGAAGGCCAGGA 2681	D	3702	GTCGCTGTGAGAGGAGACAGCTGTGGCTCAGTTCCGGGGGGAAGCATCCAAAGGCAT 3761
Q	860	uProValLYSserGluCysThrGluGluAlaGluGluGlyProAlaLYSlyLYSAspAl 880	Q	1212	eProSerThrArgValProSerAspSerAlaIleThrTYrArgGlySerIleThrIleArg 1232
D	2682	CCCCGCTCAAGAGGAGGAGGAGGAGGAGCCGAGAGGAGGCGCGCCCAAGGGCAAGAGCGC 2741	D	3762	TCCACACACACGGGTCCCTCGGACAGCGCATCAATACCGGGCTCCATCCACCGG 3821
Q	880	agIuAlaAlaGluAlaThrAlaGluGlyAlaLeuLYSAlaGluLYSlyLYSly 900	Q	1232	YThrProAlaAspValLeuLYSlyLYSlyIleThrArgIleIleGlyGluAspSerPr 1252
D	2742	GAGAGCGCTGAGAGCCACGCGCGAGAGGCGCTCAAGCGACAGAGAGAGAGAGCGGGAG 2801	D	3822	CAGCGACGTGACGTCTGTATCAAGGGCAACATCACAGAGATATCGGAGAGACAGCCC 3881
Q	900	rgLYArgAlaThrThrAlaLYSserSerGlyAlaProGluAspSerAspSerAlaThr 920	Q	1252	oSerArgLeuAspArgGlyArgGluAspSerLeuProLYSlyIleValIleTYrGluGlu 1272
			D	3882	GAGTCCGCTTGACACCGGCGCGGAGAGACGCTGCCCAAGGGCACGTATCTACGAAGG 3941

1272 YLYSLYSGIYHISVALLEUSERTYRGLUGLYMETSERVALTHRGLNYSERYSGI 1292
DB CAAGAAAGGCGCAAGTCTTCTCTATGAGGGGTGCATGTCTGTGACCCAGAGCTCCAAAGGA 4001
QY 1292 UASPGIYVAGSERSESERGILYPROBOMHISGLUTHRALAALAPROLYAHTHRYTAS 1312
DB 4002 GGACCGGCAAGACAGCTTCAGAGACCCCTCATGAGAGGGCGCCCGCAACGCGACCTATGA 4061
QY 1312 PHEMETGLUGIYATRGVALGIYARGALALLESERSEVALSERILEGLUGIYLEUMETGI 1332
DB 4062 CATGATGAGGGGCGGCTGGGAGAGGCCATCTCTCTACGACGATGAAAGTCTCATGGG 4121
QY 1332 YARGALALIEPROGLUARGHISERPROHISHISLEULYSGIUGIYHISHISILEAR 1352
DB 4122 CCGTGCATCCCGCGGAGACAGACGCCCCCACTCAAGAGACAGACACATCCG 4181
QY 1352 GGLYSERILETHRGINGLYIIEPROAGSERTYRVALAGIYAGINGIYUASPTYLEUAR 1372
DB 4182 CCGGTCCATCACAAAGGATCCCTCGGTCTTACGTGAGGCAACAGAGAGACTACCTGCG 4241
QY 1372 GARGIYUAIYLSLEULYVARGIYUGIYTHRPROPROPROPROPROPROSERARGAS 1392
DB 4242 TCGGAGGCGCAAGCTCTTAAAGGAGGAGGACGCTCCGCCCCCACTCCCTCACGGGA 4301
QY 1392 PLEUTHRGUAIATYRYSYTHRGIMALEUGIYPROLEULYLSLEULYSPROHIAHISGI 1412
DB 4302 CCTGACCGAGGCTTCAAGACGACGAGGCGCTGGGCGCCCTGAAGCTGAAGCGGCGCATGA 4361
QY 1412 UGLIYLEUVALIATHVALYGLUAIAGIYATGSERTIIEHISGLIUIIEPROARGIUGI 1432
DB 4362 GGGCGCTGGTGGCGCCGTGAAGAGGCGGCGCTCATCAAGAAATCCCGGACAGGA 4421
QY 1432 ULEUARGHISYTHRPROGLULEUPROLEUUALAPROAGPROLEULYSGIUGIYSERTIETH 1452
DB 4422 GCTGGCGCACAGCGCCGAGCTGCTTGGCGCCCGCGGCTCAAGAGGGCTCCATCAC 4481
QY 1452 TGINGLYTHRPROLEULYSETYRAPHYTHRGIALASERTHRYGISERTYLSYSHISAS 1472
DB 4482 GGAGGGCACCCGCTCAAGTACGACACCGGCGGCTCCACACAGCTGCCMAAAGACACA 4541
QY 1472 PYALAGSERLEULIEGLYSEPROGLIYATGTHRPHETPROVALIHISPROLEUASPYA 1492
DB 4542 CDTACCTCTCTCATGTGCGAGCCCGCGGAGGTGTCCACCCGTCACCCGCTGATGT 4601
QY 1492 IMETALASAPALASRGALILEUGIYUARGIACYSYRGLUGIYSERTYLSYSESERARPR 1512
DB 4602 GATGGCGGACGCGCGGACCTGGAAGCTGCTCTCGAGAGAGGCTGTAAGAGCGGCGC 4661
QY 1512 OGITYHTRALASERSESERGILYGLYSERTIIEALATRGIYALAPROVALIIEVALIPROGI 1532
DB 4662 AAGGACCGCCAGAGCTCGGGGGGCTCCATTGGCGGCGGCGCGGTCTGTGCTGA 4721
QY 1532 ULEUGIYVPROARGINSEPROLEUTHRYTGLIUAERPHISGLIYALAPROPHALAGI 1552
DB 4722 GCTGGGTAAAGCCGCGGAGAGCCCTGTGACTATGAGACCAAGGGGCAACCTTTGCCGG 4781
QY 1552 YHISLEUPROARGIYSEPROVALIYHMETARGIYUPROTHYPROARGLEUNGILUGI 1572
DB 4782 CCACTCCCAAGAGTTCGCGCTGTGACATGCGGAGGCCACCGCGCTGTGAGAGGG 4841
QY 1572 YSERLEUSERSESERIYALASERGINAPARGIYLSLEUTHRYTHRYTHRYTHRYTHRYTH 1592
DB 4842 CAGCCTTTCTGTCCAGCAAGCATCCAGAGCCGAAAGCTGACGTGACGCTGTGAGAT 4901
QY 1592 EALYVSESPROHISERTHRYALPROGIYHISHISPROHISERPROIIESEPROTYRGI 1612
DB 4902 CCGCAAGTCCCGCACAGACCGTGGCCGAGCACCAACCAACCATCTCGGCTCATGA 4961
QY 1612 UHISLEUARGIYVALSERGIYVALASPLEUTYRARGSEHISILEPROLEUALAP 1632
DB 4962 GCACTGCTTCGGGGGCTGAGTGGCTGAACTGTATCGAGGACATCCCTTGCGCTT 5021

1632 EASPPROTHSERIIEPROARGIYIIEPROLEUASAPALAAALATYRTYLEUPR 1652
DB 5022 CCAACCCCACTCATATACCCCGGCGCATCTCTGAGCGAGCGGCTGCTTACTACTGCGC 5081
QY 1652 OARGHISLEUALAPROAENPROTHRYTYPROHISLEUTYRPROPTYRILEULIEARGSI 1672
DB 5082 CCGACACCTGGCCCGCAACCCCACTTACCGGACCTGTACCACTTACTACTACTCGCGG 5141
QY 1672 YTYRPROAPTHRALAALILEUGIYUASNARGIINTHRIIEIASNAPTYRIETHISE 1692
DB 5142 CTACCCCGACAGGGGGCGCTGAGAACCGGACAGCATCATCAATGACTATCATCTCTC 5201
QY 1692 TGINGLIMETHISHISAEHTHRYALATHRALAMETALAGIYAGIYALASPMETLEUARGI 1712
DB 5202 GCAGAGATGACCAACAAACGCGCACCGCATGGCCGACGAGCTGATATCTGAGGGG 5261
QY 1712 YLEUSERPROARGIYUSERSESERLEUALALEUANTYRZALAGIYPROARGIYILEI 1732
DB 5262 CTTCTGGCCCGGAGTCTCGCTGGCACTCAACTHAGCTGGGGTCCCGAGGATCAT 5321
QY 1732 EASPLEUSERGINVALPROHISLEUPROVALILEUVALPROPROTHPROGIYTHRYPROA 1752
DB 5322 CGACCTGTCCAAATGCGCACACTGTGCTGTGCGGCCCGACACAGGACCCAGCG 5381
QY 1752 ATHRALAMETASAPAGLEUALATYRILEUPROTHRALAPROGINPROPHASERSETHRY 1772
DB 5382 CACCGCATGACCGGCTTGTCTTACTCTCCACCGCGCCCGACCTTACAGACCGGCA 5441
QY 1772 ASERSESERPROLEUSERPROGIYGLIYPROTHRISHLEUTHRYSPROTHRYTHRYSE 1792
DB 5442 CAGCAGCTCCCACTCTCCCGAGAGTCCAAACACTTGAACAAACCAACCAACGAGT 5501
QY 1792 RSESERGIUARGIYUARGAPARGAPARGIYUARGAPARGAPARGIYUARGIULY 1812
DB 5502 CTGCTCGAGCGGAGCGAGACCGGATGAGAGCGGAGCGGATCGGAGCGGAGAAA 5561
QY 1812 SERTILEUTHSERTHRYTHRYVALGIYHISHISAPROLIETTPARGPROGIYTHRYGI 1832
DB 5562 GTCCATCTCCACAGTCCACACAGCGTGGAGACGACCACTGTGAGCTGTGACAA 5621
QY 1832 UGINSERSEGIYSESERGILYSESERGILYGLIYGLIYGLIYSESERSEARARGPROA 1852
DB 5622 GCAGAGACCGGACGAGCGAGCGAGCGAGCGGCGGGGTGGGGGACGACGACCGCCCGC 5681
QY 1852 ASERHISERHISALAHISGLIYHISERPROIIESEPROARGTHRYNAPALALEUGI 1872
DB 5682 CTCCACCTCCAGTGGCCACAGACACTGCCCATCTCCCTCGGAGCCAGAGATGCCCTCA 5741
QY 1872 NGIUDARGPROSERVALLEUHHISAEHTHRYIIMETLYSGIYIIEIETHRYALVALIUIPR 1892
DB 5742 GCAGAGACCAAGTGTGCTTCAACAACACAGGCAATGAAGGTATCATACCGCTGTGAGCC 5801
QY 1892 OSERTYSPROTHRYVALLEUARGSETHRYTHRYTHRYTHRYTHRYTHRYTHRYTHRYTH 1912
DB 5802 CAGCAAGCCACGAGTCTTAGAGTCACTTCACTCTCTCAACCGTTCCGCGACGCTGCCAC 5861
QY 1912 RPHEPROPROALATHRHHISYAPROLEUGIYGLIYTHRIEUAERGIYVALIYTYRPROTHRIE 1932
DB 5862 ATTTCCACTGCAACCACTGCCACTGGGCGGACCTCGATGGGGGTCTACCTTACCTCT 5921
QY 1932 UMETGIUPROVALILEUENUPOLYSGIYUAPROARGVALIYALATGPPROGIYUARGPROAR 1952
DB 5922 CATGGAGCCGCTTGTGCTTCCAAAGAGCCCGCGGCTGGCGCGGACGAGCGGCGCG 5981
QY 1952 GALIASPTHRGIYHISALAPHELEUALAYSPROPROALATGSESGIYLEUGIUPROA 1972
DB 5982 AGCAGACACCGGCAATGCTTCTGCGCAAGGCCCAAGCCGCTCGGGCTGGAAGCCCGC 6041
QY 1972 ASERSESPROSERTYSGIYSEGIYUPROARGPROLEUVALIYPROVALISERGILYHISAL 1992
DB 6042 CTCCTTCCCGACAGAGGCTCGGAGGCGCGGCGCTTGAAGCTCTGTCTGTGGCAGCGC 6101
QY 1992 ATHRIIEALATGTHRPROALALYASMLEUALAPROHISHISALASERPROASPPROPR 2012

Db 6102 CACCATGCCCCGACCCCTGGAGAAAGACCTCGACACTCACACCCGACCCGACCCGCC 6161
Qy 2012 oAlaProProAlaSerAlaSerAlaSerProHiAaRGluLysThcGlnSerLysProPheSe 2032
Db 6162 GGGCCCACTGGCTCTGGCTCTGGACCCGACCCGGGAAAAGACTAAAGTAAACCTTTTC 6221
Qy 2032 rIleGlnGluLeuGluLeuLeuSerLeuGlyTyHISGlySerSerTySerProGluG 2052
Db 6222 CATCCAGGAAGTGAAGCTCCGTTCTCTGGGTTTCAAGGAGGAGCTACAGCCCCGAAAG 6281
Qy 2052 yValGluProValaSerProValaSerSerProSerLeuThHiAaPlySGlyLeuProLy 2072
Db 6282 GGTGGAGCCCGCTGACCCCTGGAGCTCACAGCTCAACCCAGCAAGGGGCTCCCCAA 6341
Qy 2072 sHISleuGluGluLeuAaPlySerThISleuGluGlyLeuLeuAaRProLysGlnProG 2092
Db 6342 GCACTGGAAAGAGTCCAGCAAGAGCCACTGGAGGGGAGCTGGGCCCAAGACGCAAG 6401
Qy 2092 yProValaLysLeuGlyGlyGluAaAaHISleuProHiAaLeuAaRProLeuProGluSe 2112
Db 6402 CCCCCTGAAGCTTGGCGGAGAGCGCCGACCTCCACCTCGGCGCCGCTGCTGAGAG 6461
Qy 2112 rGlnProSerSerSerProLeuLeuGlnThraAaProGlyValaLysGlyHISGlnAaRyVa 2132
Db 6462 CCAAGCCCTGCTCCAGCCGCTGCTCCAGACGCGCCCAAGGGGTCAAGGTCAACGCGG 6521
Qy 2132 lValThLeuAaGlnHISleuSerGluValaHISleuThAaPlyTyThraThraHISleu 2152
Db 6522 GGTCACTTGAAGCCAGCAATCAAGTGAAGTATCAACAGAGTCAACCCGCGCACCC 6581
Qy 2152 oGlnGlnLeuSerAlaProLeuProAlaProLeuTySerPheProGlyAlaSerCysPr 2172
Db 6582 ACAAGCACTAGCGGCACTCCCTGCCGCCCTTACTCTCTCTGGGCGACCTGCC 6641
Qy 2172 oValLeuAaPlyAaRgRProProSerAaPlyLeuTyLeuProProAaPlyHISGlyAl 2192
Db 6642 CGTCTGGAAGCTCCGCGCGCCACCAAGTGAAGTCTTACTCCCGCCCCGAGCATGTGC 6701
Qy 2192 aProAlaAaRgLySerProHiAaSerGlyGlyTyLeuAaRgSerProGluProAaPlyTh 2212
Db 6702 CCGGCGCCGCTGGCTCCCGCCAGCGAAGGGGGGMAAGGTCTTCAAGCAAGCAAGAC 6761
Qy 2212 rSerValaLeuGlyGlyGlyGluAaPlyIleGluProValaSerProProGluGlyMetTh 2232
Db 6762 GTGGGTCTTGGGTGGTGGTGAAGCAAGTATTGAACCTGTCTCCACCGAAGGGCATGAC 6821
Qy 2232 rGluProGlyHISerAaRgSerAlaValTyProLeuLeuTyAaRgAaPlyGlyGlnTh 2252
Db 6822 GAGAGCCAGGCACTCCCGGAGTGTGTGATCCGCTGTATCCGGGATGGGAAACAGAC 6881
Qy 2252 rGluProSerAaRgMerGlySerTySerProGlyAaRThSerGlnProProAlaPhePh 2272
Db 6882 GAGGCCAGAGAGATGGGCTTCAAGTCTCAAGGCAACAGCAGCCGCGACCTTTCTT 6941
Qy 2272 eSerTyLeuThGlySerAaRgSerAlaMetValTySerTyLeuGlnGlnIleAaPly 2292
Db 6942 CAGCAAGCTAGCCGAGAGCAATCCGCAATGTCMAAGTCCAAGAGCAAGATCAACAA 7001
Qy 2292 sTyLeuAaPlyThHISaAaRgAaGlnGluProGlyTyAaPlyISerGlnProGlyThrG 2312
Db 7002 GAAGCTGAACACCAACCAAGATGAGCTGAATACATATCAGCAGCCTGGAGCGA 7061
Qy 2312 uIlePheAaPlyProAlaIleThrGlyThrGlyLeuMetThTyAaRgSerGlnAlaVa 2332
Db 7062 GATCTTCAATATATCCGCGCATCAACCGAAGAGGCTTATATACCTTATGAAGCGAGG 7121
Qy 2332 lGlnGlnHISleuSerThraAaPlyGlyLeuGluAaIleIleAaRgTyAaLeuMetG 2352
Db 7122 GCAAGAAACATGCAAGCAACATGAGGCTGAGAGCAATATTAAGAAAGCACTCATGG 7181
Qy 2352 yTyTyTyAaPlyGlnTyProGlyLeuSerProProLeuSerAlaAaPlyAaPlyLeuAa 2372

Db 7182 TAAATATGACAGTGGAGAGATGCCCGCTCAGGCCCAATGCTTTTAACCTCTGAA 7241
Qy 2372 nAlaSerAlaSerLeuProAlaAlaMetProIleThraAlaAaPlyAaRgSerAaPly 2392
Db 7242 TGGCAGTGGCAAGCTTGGCTGCTGTATGCTTATGCCATTAACCTGTGACGAGAGTACCA 7301
Qy 2392 sThrLeuThSerProGlyGlyGlyGlyAaAaLysValaLysSerGlyAaRgProSerSerA 2412
Db 7302 CACACTCACTCCGCAAGTGGCGGAGAGCCCAAGGTCTGTGGCAGACCCAGCAGCCG 7361
Qy 2412 GlyValaLysSerProAlaProGlyLeuAaLysGlyAaPlyAaRgProProSerValaSer 2432
Db 7362 AAAAGCAAGTCCCGGCGCCGCGGCTGCGCATCTGGGAGCCGACCTCTGTCTCTC 7421
Qy 2432 rValHISerGlyGlyAaPlyAaAaRgAaGlyThrProLeuThAaAaRgValTyProGluA 2452
Db 7422 AGTGAAGTGGAGAGAGTGAAGTGAAGCCGCGAGCGGCTTACCAACCCGCTGTGGAGGA 7481
Qy 2452 PatGProSerSerAlaGlySerThraProPheProTyAaPlyProLeuIleMetAaRgLeuG 2472
Db 7482 CAGGCTTGTCCGCAAGTTCAGAGCATTCCTCTCAACCCCTGATCATGGGCTGCA 7541
Qy 2472 nAlaGlyValaMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeuAl 2492
Db 7542 GCGGGGTGTATGCTTCCCGCACCCGCGGCTTCCCGCGGCGAGCGGCGCCCTGCG 7601
Qy 2492 aGlyProHISleuAaPlyAaPlyGluProPlyAaPlyProLeuCysSerGlnTyGlyTh 2512
Db 7602 TGGCCCCACCAAGCTGGGAGCAAGAGCCCAAGCACTGTCTGTGCGAGTACGAGAC 7661
Qy 2512 rLeuSerAaPlySerGlu 2517
Db 7662 ACTCTCGAAGCGAG 7677

RESULT 8
US-10-723-860-1739
; Sequence 1739, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882, 0193, NPLS01
; CURRENT APPLICATION NUMBER: US/10/723, 860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429, 739
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1739
; LENGTH: 8686
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-1739

Alignment Scores:
Pred. No.: 0 Length: 8666
Score: 12978.00 Matches: 2487
Percent Similarity: 98.57% Conservative: 3
Best Local Similarity: 98.46% Mismatches: 9
Query Match: 98.21% Indels: 28
DB: 21 Gaps: 3

US-09-522-753-5 (1-2517) x US-10-723-860-1739 (1-8686)
Qy 1 MetSerGlySerThrGlnLeuValaGlnThraAaRgAlaThrGluProAaRgTyPro 20
Db 157 ATGTCGGGCTCCACAGAGCTGTGGCAAGAGTGAAGGAGGCACTAGGCGGCTACCGG 216
Qy 21 ProHISerLeuSerTyProValGlnIleAaRgThraHISleuThraPlyAaGlyLeuLeu 40

Db	217	CCCCAGACCTTCTCTACCAGTGCAGATCGCCCGGAGCAGACACGAGACTGGGCTCTG	276
Oy	41	GlutryGlnHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln	60
Db	277	GAGTACCGAGCAACACTCCCGCAGCATATCTCTCCCACTGTGCGCCGGCTCATATCCAG	336
Oy	61	ProGlnArgArgProSerLeuLeuSerGlnPheGlnProGlyAsnGlnLysSerGln	80
Db	337	CCCCAGCGCGGAGGCGCTCTCTGTCGTAGTTCAGCCCGGGAATGAACGGTCCAG	396
Oy	81	GlueuHisLeuArgProGlnSerHisSerTyrLeuProGlnLeuGlyLysSerGlnMet	100
Db	397	GAGCTCCACCTGCGGCGCAGAGTCCCACTCATACCTGCCGAGACTGGGAAATCAGAGTGG	456
Oy	101	GluPheIleGlnSerLysArgProArgLeuGlnLeuLeuProAspProLeuLeuArgPro	120
Db	457	GAGTTCATTGAAGCAAGCGCCCTCGGTAGAGCTGTGCTGACCCCTGCTGGACGG	516
Oy	121	SerProLeuLeuAlaThrGlyGlnProAlaGlySerGlnAspLeuThrLysAspArgSer	140
Db	517	TCACCCCTGCTGGCCACGCGGCGAGCTCGGAGATCTGAAGACTCACCAGACCGTGAAC	576
Oy	141	LeuThrGlnLysLeuGlnProValSerProProSerProProHisThrAspProGlnLeu	160
Db	577	CTGACGGGGAAGCTGGAACCGGTGTCTCCCGCCACCCCGGCAACTGACCTTGAGCTGG	636
Oy	161	GluLeuValProProArgLeuSerLysGlnGlnLeuIleGlnAsnMetAspArgValAsp	180
Db	637	GAGCTGGTGGCCGCAACGGCTGTCCAAAGAGGAGCTGATCCAGAAACATGGAACCGGTGAC	696
Oy	181	ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeu	200
Db	697	CGAGAGATCACCATGTGTAGAGAGCAGATCTCTAAGCTGAAGAGAAACAGACACAGCTG	756
Oy	201	GluGlnGlnAlaAlaLysProProGlnProGlnLysProValSerProProProIleGln	220
Db	757	GAGGAGAGGCTGCCAAGCGCCGCGAGCTTGAGAGCCCGTGTACCGCCGCTCATCGAG	816
Oy	221	SerLysHisArgSerLeuValGlnIleIleTyrAspGlnAsnArgLysLeuAlaGlnAla	240
Db	817	TCGAAGCACCGCAGCTGTGTGCAGATCATCTACAGCAGAAACCGGAAAGAGCTGAACCT	876
Oy	241	AlaHisArgIleLeuGlnGlyLeuGlnLysProGlnValGluLeuProLeuTyrAsnGlnPro	260
Db	877	GCACATCGGATTTCTGGAAGGCGCTGGGGCCCCGAGTGGAGACTGCGCTGTACAAACACGCC	936
Oy	261	SerAspThrArgGlnTyrHisGlnAsnIleLysIleAsnGlnAlaMetArgLysLysLeu	280
Db	937	TCCGACACCCCGCAGTATCATAGAACATCAAAATAAACAGAGCCATCGGAAGAAAGCTA	996
Oy	281	IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGln	300
Db	997	ATCTTGTACTTCAAGAGAGAAATACCGCTCGGAAACATGGAGACAAAGTTCTGCCAG	1056
Oy	301	ArgTyrAspGlnLeuMetGlnAlaLeuGlnLysLysValGluArgIleGlnAsnAsnPro	320
Db	1057	CGCTATGACCAAGCTCATGAGAGGCGCTGGGAGAAAGAGTGGAGCGCATCGAAGAACACCC	1116
Oy	321	ArgArgArgAlaLysLysGlnSerLysValArgGlnTyrTyrGlnLysGlnPheProGlnIle	340
Db	1117	CGGCGCGCGGCGCAGAGAGCAAGGTTGCGCGATCTACGAAACCAATGTTCCCTGAGATC	1176
Oy	341	ArgLysGlnArgGlnLeuGlnGlnLysArgMetGlnSerArgValGlyArgGlySerGly	360
Db	1177	CGCAAGCACCGCGAGCTGCAAGAGCGCATTCAG--AAGGTGGGCCAAGCGGGCAGTGGG	1233
Oy	361	LeuSerMetSerAlaAlaArgSerGlnHisGluValSerGlnIleLeuAspLysLeuSer	380
Db	1234	CTGTCCATGTGCGCCCGCGCAGCAGACACGAGTGTCAAGATCATCGATGACCTCTCA	1293
Oy	381	GluGlnGlnLysLeuGlnLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyr	400
Db	1294	GAGCAGGAGAACCTGGAGAAAGAGATGGCGCAGCTGGCGCGGTATCCGCCCATGTGTTAC	1353

QY	401	Asp1aa	AspGlnGlnIa	Arg1Ile	Leu	SerPhe	Ile	Asn	Met	Asn	Gly	Leu	Met	Ala	Asp	Prom	Met	420				
Db	1354	GACGCTGAC	CACAGCAGCGCAT	CAAGTTCAT	CAACATCA	ACGAGGCTT	ATGAGCGGAC	CCCATG										1413				
OY	421	Lys	Val	Tyr	Leu	Asp	Arg	Gln	Val	Met	Asn	Met	Trp	Ser	Glu	Gln	Lys	Arg	440			
Db	1414	AAAGTGTC	CAAAAGACCGCC	AGGTCA	TGAACT	ATGGAGT	GTGAGCAG	GAAGAAG	CACTTC									1473				
OY	441	Arg	Glu	Lys	Phe	Met	Gln	His	Pro	Lys	Asn	Phe	Gly	Leu	Ile	Ala	Ser	Phe	Leu	Arg	460	
Db	1474	CGGAGGA	AGTTCA	TGACAG	CAATCC	CAAGAACT	TTGGCTT	ATGCGAT	CGCATTC	CTCC	GTGAGAGG									1533		
OY	461	Lys	Thr	Val	Ala	Glu	Cys	Val	Leu	Tyr	Tyr	Tyr	Leu	Thr	Gly	Ser	Asn	Glu	Asn	Tyr	Lys	480
Db	1534	AAGACAG	TGGCTG	AGTGGCT	CTCTCT	ATTACT	CTAC	TGACT	TAAGAA	GAATGA	GAAC	TTAA	G								1593	
OY	481	Ser	Leu	Val	Arg	Arg	Ser	Tyr	Arg	Arg	Arg	Gly	Lys	Ser	Gln	Gln	Gln	Gln	Gln		500	
Db	1594	AGCCTGT	GTGAGACG	AGAGTAT	CGGCGCC	CGGCAAG	ACAGC	AGCAGC	AGCAAC	AAACAG	CAGCAG										1653	
OY	501	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	520	
Db	1654	CAGCAGC	AGCAGC	AGCAGC	AGCAGC	AGCAGC	AGCAGC	AGCAGC	AGCAGC	AGCAGC	AGCAGC	AGCAGC	AGCAGC	AGCAGC	AGCAGC	AGCAGC	AGCAGC	AGCAGC	AGCAGC	AGCAGC	1713	
OY	521	Asp	Glu	Lys	Glu	Lys	Glu	Lys	Glu	Asn	Glu	Asn	Glu	Asn	Glu	Asn	Glu	Asn	Glu	Asn	540	
Db	1714	GATGAG	AAAGAA	GAAAGAA	AAAGGAG	GGCGGAG	AAAGAA	GAAAGAA	GAAAGAA	AAAGGAG	GGCGGAG	AAAGAA	GAAAGAA	GAAAGAA	AAAGGAG	GGCGGAG	AAAGAA	GAAAGAA	GAAAGAA	GAAAGAA	1773	
OY	541	Asp	Lys	Glu	Asp	Leu	Leu	Lys	Glu	Ser	Thr	Asp	Asp	Thr	Ser	Gly	Lys	Asp	Asn	Asp	Glu	560
Db	1774	GACAA	GAGAA	GACCTCT	CTCAAG	GAGAA	GAGAA	GAGAA	GAGAA	GAGAA	GAGAA	GAGAA	GAGAA	GAGAA	GAGAA	GAGAA	GAGAA	GAGAA	GAGAA	GAGAA	1833	
OY	561	Lys	Glu	Ala	Val	Ala	Ser	Lys	Gly	Arg	Lys	Thr	Ala	Asn	Ser	Gln	Lys	Arg	Tyr	Gly	580	
Db	1834	AAAGAG	CGCTGT	GGCCTC	CAAAAG	CGCCCA	AAAC	CTGCCA	ACAGCC	AGGAA	AGACCC	CAAA	GCC								1893	
OY	581	Arg	Ile	Thr	Arg	Ser	Met	Ala	Asn	Glu	Ala	Asn	Ser	Glu	Glu	Ala	Ile	Thr	Pro	Gln	600	
Db	1894	CGCAT	CACCCCG	CTCAAT	GCTAAT	GAGGCA	CAGGAG	AGGCA	TCAC	CCCC	CAGCAG										1953	
OY	601	Ser	Ala	Glu	Leu	Ala	Ser	Met	Glu	Asn	Asn	Glu	Ser	Arg	Thr	Arg	Glu	Glu	Met		620	
Db	1954	ACGCG	CGAGCT	GGCCTC	CAATGA	AGCTGA	TGAAGAT	TTCCGCT	GGACAG	AAAGAA	TAATG										2013	
OY	621	Glu	Thr	Ala	Lys	Lys	Gly	Leu	Leu	Gln	His	Gly	Arg	Asn	Trp	Ser	Ala	Ile	Ala	Arg	Met	640
Db	2014	GAAAC	AGCCAA	GAAAGAT	CTCTCG	GAAACAG	CGGCGCA	AC	TCGGC	CAATCG	CCCGCAT	CGCCG	GATG								2073	
OY	641	Val	Gly	Ser	Lys	Thr	Val	Ser	Gln	Cys	Lys	Asn	Phe	Tyr	Phe	Asn	Tyr	Lys	Val	Arg	Gln	660
Db	2074	GTGGG	CTCCAA	ACGTGT	GTGCGAG	GTGAAGAACT	TTACTT	CAACT	CAACAA	GAAAGAG	CAG										2133	
OY	661	Asn	Leu	Asp	Glu	Ile	Leu	Gln	His	Lys	Lys	Leu	Lys	Met	Glu	Lys	Glu	Arg	Asn	Ala	Arg	680
Db	2134	AACTC	CGATGA	ATCTTG	CAAGCAG	CAACAAG	CGTGA	ATGA	TGAAGA	AGAGAG	AGAA	CGCGCG									2193	
OY	681	Arg	Lys	Lys	Lys	Val	Ala	Pro	Ala	Ala	Ala	Ser	Glu	Glu	Ala	Ala	Phe	Pro	Pro	Val	700	
Db	2194	AGGAAGA	AGAAAGAA	CGCGCG	CGCGCG	CGCGCG	CGAGAG	AGAGCG	CTGCAT	TCGCCCG	CGCGGTG										2253	
OY	701	Glu	Asp	Glu	Glu	Met	Gln	Lys	Ser	Gly	Val	Ser	Gly	Ala	Asn	Gln	Glu	Glu	Met	Val	Glu	720
Db	2254	GAGGAT	AGAGAT	GTGAGG	CGTGGCG	CGTGA	CCGGA	AAATGA	GAGAGAT	GTGTG	AGAG										2313	
OY	721	Ala	Glu	Ala	Leu	Ile	Ala	Ser	Gly	Asn	Glu	Val	Pro	Arg	Gly	Glu	Cys	Ser	Gly	Pro	Ala	740
Db	2314	GCTGAA	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	2332	
OY	741	Thr	Val	Asn	Asn	Ser	Ser	Asp	Thr	Glu	Ser	Ile	Pro	Ser	Pro	His	Thr	Glu	Ala	Ala	Lys	760
Db	2334	ACTGTCA	ACAAAG	CGCTCA	GACAC	CCGAGAC	ATCTCC	CTCTCT	CTAC	ACTGAG	GGCCG	CCAG									2382	

Qy 761 AspThrGlyGlnAnGlyProLyseProProAlaThrIleuGlyValAspGlyProProPro 780
Db 2283 GACACAGGGCGAGAAATGGGGCCCAAGCCCAAGCCCTGGGGCGCGAGCGGGCCACCCCA 2442
Qy 781 GlyProProThrProPro-ArgArgThrSerArgAlaProIleGluProThrProAlaSe 800
Db 2443 GGGGCCACCCACCCACCCAGAGACATCCGGCC-CCCACTAGTCCACCCCGGCTC 2501
Qy 800 rGluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProProPr 820
Db 2502 TGAAGCCACTTAAGCCCTTAAGCCCCCAAGCACCCCACTTCCCTTCACTCTCC 2561
Qy 820 OValValProLyseGlyGluGlyGluGluGluThraAlaAlaAlaProProValGluGlu 840
Db 2562 TGtGTGCTCCCAAG 2621
Qy 840 YGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 860
Db 2622 GAG 2681
Qy 860 uProValLyseSerGlyCysThrGluGluGluGluGluGluGluGluGluGluGluGlu 880
Db 2682 GCCCGTCAAG 2741
Qy 880 agluAlaagluAlaThrAlaGluGlyAlaLeuLyseAlaGluLyseGlyGlySe 900
Db 2742 GAG 2801
Qy 900 rGlyArgAlaThrThrAlaLyseSerGlyAlaProGluAspSerAspSerAlaTh 920
Db 2802 CGGAG 2861
Qy 920 rCysSerAlaAspGlyValAspGluAlaGluGlyGlyAspLyseAspLyseLeuSerPr 940
Db 2862 CTGCAAGTGCAG 2921
Qy 940 oArgProSerLeuLeuThrProThrGlyAspProArgAlaAspAlaSerProGluYpR 960
Db 2922 AAGGCCAGAGCTCTCAAG 2981
Qy 960 oLeuAspLeuLyseGluLeuLyseGluGluGluAlaAlaAlaIleProProIleGluValThrLy 980
Db 2982 ACTGAGACTTGAAG 3041
Qy 980 sValHisGluProProArgGluAspAlaAlaProThrLyseProAlaProAlaProPr 1000
Db 3042 AGTTCATGAG 3101
Qy 1000 oProProGluAspLeuGluProGluAspAlaProGluGluProGlySerSerProAr 1020
Db 3102 GCCACCGCAAAACCTGAG 3161
Qy 1020 gGlyLyseSerArgSerProAlaProProAlaAspLyseGlu----- 1033
Db 3162 GGGCAG 3221
Qy 1034 -----AlaPheAlaAlaGluAlaGluGluGluGluGluGluGluGluGluGluGlu 1052
Db 3222 CCCAG 3281
Qy 1052 yLeuProPheProValProProArgGluValIleValAlaSerProHisAlaProAspPr 1072
Db 3282 CTGAGCTTCCCGTGGCCCGGTGAGGTGATCAAGGCTCCCGCATGCCCGGAGACC 3341
Qy 1072 oSerAlaPheSerTyraAlaProProGlyHisProLeuProLeuGlyLeuHisValThrAla 1092
Db 3342 CTAGAGCTTCTCAAGCTCACTGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3401
Qy 1092 aaArgProValLeuProArgProProThriIleSerAspProProProLeuIleSerSerAl 1112
Db 3402 CGGGAG 3461
Qy 1112 aLyseHisProSerValLeuGluGluGluGluGluGluGluGluGluGluGluGluGlu 1132

Db 3462 CAAG 3521
Qy 1132 mLeuHisValProLyseSerGluHisAlaLyseAlaProValGlyProValThrMetGlyLe 1152
Db 3522 GCTCAGAGTCCCGTACAG 3581
Qy 1152 uProLeuProMetAspProLyseLyseLeuAlaProPheSerGlyValLyseGluGluGlu 1172
Db 3582 GCCCTGCCATGAG 3641
Qy 1172 uSerProArgGlyGluAlaGlyProProGluSerLeuGlyValProThrAlaGluGlu 1192
Db 3642 GTCCCAAG 3701
Qy 1192 aSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLyseGly 1212
Db 3702 GTCCGTCTGAG 3761
Qy 1212 eProSerThrArgValProSerAspSerAlaIleThrTyraArgLyseIleThrHisGly 1232
Db 3762 TCCAG 3821
Qy 1232 yThrProAlaAspValLeuTyryLyseGlyThrIleThrArgIleIleGlyGluAspSerPr 1252
Db 3822 CAGCCAGCTGAG 3881
Qy 1252 oSerArgLeuAspArgGlyValArgGlyAspSerLeuProLyseGlyHisValIleTyryGlu 1272
Db 3882 GAGTCGCTTGAAG 3941
Qy 1272 yLyseGlyHisValLeuSerTyryGluGlyGlyMetSerValThrGlnCysSerLeuGly 1292
Db 3942 CAAG 4001
Qy 1292 uAspGlyArgSerSerSerGlyProProProHisGluThrAlaAlaProLyseArgThyTyra 1312
Db 4002 GAG 4061
Qy 1312 pMetMetGluGlyValTyraGlyValTyraGlyAlaIleSerSerAlaSerIleGluGlyLeuMetGly 1332
Db 4062 CATGATGAG 4121
Qy 1332 yArgAlaIleProProGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 1352
Db 4122 CCGTGCATCCCGAG 4181
Qy 1352 gGlySerIleThrGlnGlyIleProArgSerTyryValGluAlaGluGluAspTyryLeuAr 1372
Db 4182 CGGATCATCAACAAG 4241
Qy 1372 gArgGluAlaLyseLeuLeuLyseArgGlyGlyIleThrProProProProProProSerArgAs 1392
Db 4242 TCGGAG 4301
Qy 1392 pLeuThrGluAlaThrLyseThrGluAlaLeuGlyProLeuLyseLeuLyseProAlaHisGly 1412
Db 4302 CTGAGCCAG 4361
Qy 1412 uGlyLeuValAlaThrValLyseGluAlaGlyArgSerIleHisGlyIleProArgGluGly 1432
Db 4362 GGGCTGTGTGGAG 4421
Qy 1432 uLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLyseGlyGlySerIleTh 1452
Db 4422 GCTGAG 4481
Qy 1452 rGlnGlyThrProLeuLyseTyraAspThrGlyAlaSerThrThrGlySerLyseHisHisAs 1472
Db 4482 GAG 4541
Qy 1472 pValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspVal 1492

Db 4542 CGTAGCCTCCCTCATCGGAGCCCGCGCGAGCGTTCCACCCGTGACCCCGTGAATG 4601
Qy 1492 IMeTAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuYserArgPr 1512
Db 4602 GATTGGCCGAGCCCGGGACCTTGAAAGCTGCTGCTACAGAGAGAGCTTAAGAGCCGGCC 4661
Qy 1512 oGlyThrAlaSerSerSerGlyYserIleAlaArgIleAlaProValIleValProGl 1532
Db 4662 AAGGAGCCCGCAGAGCTCGGGGGGCTCCATTGGCGCGGGGCCCGGATTTGTCCTGA 4721
Qy 1532 uLeuGlyLysProArgGluSerProLeuThrTyrGluAspHisGlyAlaProPhaIleAl 1552
Db 4722 GTTGGGTAAAGCCGCGCAGAGCCCTGTGACCTAATGAGACCAAGGGGCACCTTTGCCG 4781
Qy 1552 yHisLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGluG 1572
Db 4782 CCACTTCCACAGAGTTCCGCGCTGACCATGCGGGAGCCACCGCGCTGAGAGAGG 4841
Qy 1572 ySerLeuSerSerSerLeuAlaSerGluAspArgLysLeuThrSerThrProArgGluI 1592
Db 4842 CAGCCTTTGTCAGAGAGCATCCAGAGCCGAAAGCTGACGTCGACGCTGTGAGAT 4901
Qy 1592 eAlaLysSerProHisSerThrTyrAlaProGluHisIleSProHisProIleSerProTyrG 1612
Db 4902 CGCCAAAGTCCCGCAGACAGACCGGTGCCAGACACCAACCCCATCTCGCCCTATGA 4961
Qy 1612 uHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPh 1632
Db 4962 GCACTGTCTTCGGGGGTAGTGCTGTGACCTGTATCCAGGCAATATCCCTGGCCTT 5021
Qy 1632 eAspProThrSerIleProArgGlyLysLeuAspAlaAlaAlaAlaTyrTyrLeuPr 1652
Db 5022 CCAACCCACCTCATACCCCGCGGATCCCTGTGAGCGAGCGCTGACTACTACCTGCC 5081
Qy 1652 oArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgG 1672
Db 5082 CCGACACCTGTGGCCCCCAACCCCACTACCGGACCTGTACCCACTTACTTCGCGG 5141
Qy 1672 yTyrProAspThrAlaAlaLeuGluAsnArgGluThrIleIleAsnAspTyrIleThrSe 1692
Db 5142 CTACCCCGACCGCGCGCTGAGAAACCGGCAAGCATCATATATAGATCACTACCTC 5201
Qy 1692 rGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgG 1712
Db 5202 GAGAGCAGATGACACACAAACCGGCCACCGCATGGCCAGCGAGCTGATGTGAGGG 5261
Qy 1712 yLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyLysI 1732
Db 5262 CCTCTGCGCCCGGAGTCTCTGTGGACATCACTACGTGCGGGTCCCGAGGACATCAT 5321
Qy 1732 eAspLeuSerGluValProHisLeuProValLeuValProProThrProGlyThrProAl 1752
Db 5322 CCAACTGTCCCAAGTGCACACTGTGCTGTGCTGCTGCCCCCGACACACAGGACCCCA 5381
Qy 1752 aThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheserSerArgH 1772
Db 5382 CACCGGCATGAGACCGCTTGCTTACCTCCCAACCGGCCCCAGCCCTTGACAGCGGCC 5441
Qy 1772 sSerSerSerProLeuSerProGlyGlyProThrHisIleLeuThrLysProThrThrSe 1792
Db 5442 CAGCAGCTCCCACTCTCTCCAGAGGTCCAAACACTTGAACAAACCAACCAACCAAGTC 5501
Qy 1792 rSerSerGluArgGluArgAspArgAspArgLysArgAspArgAspArgGluArgGlu 1812
Db 5502 CTGTCTCGAGCGGAGAGAGAGCCGGATCGAGAGCGGAGCCGGATCGGAGCGGAGAAA 5561
Qy 1812 sSerIleLeuThrSerThrThrThrValGluHisAlaProIleTyrArgProGlyThrG 1832
Db 5562 GTCCATCTCTACATCCACCAAGCGGTGAGACAGCAACCATCTGAGAGACTGTGTACAG 5621
Qy 1832 uGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProAl 1852
Db 5622 GCAGAGCAGCGGAGCAGCGGAGCGGCGGGGTGGGGGAGCAGCAGCGCGCCCGCC 5681

Qy 1852 sSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGluAspAlaLeuG 1872
Db 5682 CTCCCACTCCCATGCCACACAGACTCGGCCCATCTCCCTCGGAGCCAGAGATGCCCTTCA 5741
Qy 1872 nGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluPr 1892
Db 5742 GCAGAGACCCCAAGTGTCTTCAACAAACAGGCAATGAAGGATATCATACCGCTGTGAGCC 5801
Qy 1892 oSerLysProThrValLeuArgSerThrSerHisSerProValArgProAlaAlaTh 1912
Db 5802 CAGCAAGCCCAAGTCTCTGAGGTCCACTCCACTCTCTACCCGTTGCGCCAGCTGCCAC 5861
Qy 1912 rPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLe 1932
Db 5862 ATTTCCCACTGCCACCACTGCCCACTGGCGGAGCCCTCGATGGGGTCTTACCTTACCT 5921
Qy 1932 uMetGluProValIleLeuLeuProLysGluAlaProArgValAlaArgProGluArgProAr 1952
Db 5922 CATGAGCCCGTCTTGCTGCGCAAGAGGCCCCCGGGTCCCGGCCAGAGCGGCCCG 5981
Qy 1952 gAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAl 1972
Db 5982 AGCAGACACCGGCATGCTTCTCGCCAAAGCCCGACCGCTCGGGCTGAGGCCCGC 6041
Qy 1972 sSerSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAl 1992
Db 6042 CTCCTCCCCCAAGAGGCTCGAGGCCCGGCCCTAGTGCCTCTGTCTGGCCACGC 6101
Qy 1992 aThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProPr 2012
Db 6102 CACCATGCGCCGACCCCTGCGAAGAACTCGACATCCACACCGCCAGCCGAGCCCGCC 6161
Qy 2012 oAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPhese 2032
Db 6162 GCGGCACTGTGCTCGGCTCGGACCCGACCGGAGAAAGACTCAAAAGTAAACCTTTTC 6221
Qy 2032 rIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluG 2052
Db 6222 CATCCAGAACTGGAATCGTTCTCTGGGTTACCAACGACAGCTACAGCCCCCAAG 6281
Qy 2052 yValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLy 2072
Db 6282 GGTGAGCCCGTCAAGCCCTGTGACTCACAGTGTGACCCACGACAAGGGGCTCCCA 6341
Qy 2072 sHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGl 2092
Db 6342 GCACTGGAAGCTGACAAAGACCACTTGAGGGGAGCTGGCCCAAGCAGCCAG 6401
Qy 2092 yProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSe 2112
Db 6402 CCCCCTGAAGCTTGGGGGAGAGCGGCCCACTCCACACTTGGCGCCGCTGCTGAGAG 6461
Qy 2112 rGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgVa 2132
Db 6462 CAGGCCCTGTTCAGCCCGCTGCTCCAGACCGGCCAGGGGTCAAAGGTCACCAAGGGGT 6521
Qy 2132 lValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisI 2152
Db 6522 GGTCACTGTGGCCCAAGCATCATGAGGTCAATCACAGAGACTACACCGGACCAACCC 6581
Qy 2152 oGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysPr 2172
Db 6582 ACAGCAGCTCAAGGACCCCTGCGCCCGCCCTCTCTCTCTCCCTGGGGGCCAGCTGCCC 6641
Qy 2172 oValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHisGlyAl 2192
Db 6642 CGTCTGAGACTTCGCGCGCCACCAAGTACTTACCTTCGCGCCCGGACCATGCTGC 6701
Qy 2192 aProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnLysTh 2212
Db 6702 CCGGGCCGTGTCTCCCCCAAGCAGCAAGGGGCAAGAGGTCTCCAGAGCCAAACAAGAC 6761

QY 2212 rSerValleuGlYglYglYuaPglYleGluproValSerProPogluJymetTh 2232
Db 6762 GTCCGCTTGGGTGGTGGTGAAGACGATTAACCTGTGTCACCGAGGGCATTGAC 6821
QY 2232 rGluproGlYhiSerTargSerAlaValTYrProleuLeuTYrTargApGlygluJmTh 2252
Db 6822 GGAGCCAGGGACTCCCGAGATGCTGTGTACCCGCTGTGTAACGGGATGGGAAACAGAC 6881
QY 2252 rGluproSerTargMetGlySerTylSerProGlYASnThSerGlnProProAlaPhePh 2272
Db 6882 GGAGCCAGAGATGGGCTTCAAGTCTCCAGGCAACCGACCGGCGGCTTCTT 6941
QY 2272 eSerTylLeuThrGlYSerAnSerAlaMetValYleSerTylSerGlnJmleAnTy 2292
Db 6942 CAGCAAGCTAGCCAGAGCAACTCCGCTATGTAAGTCAAGAGCAAGAGATCAACAA 7001
QY 2292 glybLeuAnThThiASnTargAnGluproGlYTYrASnTleSerGlnProGlYThrGl 2312
Db 7002 GAAGCTGAACACCCACCAACCGGAATGAGCTGAAATACAAATATCAGCCAGCTGGAGCGGA 7061
QY 2312 uilePheASnMetProAlaIleThrGlYThrGlYleuMetThrTYrTargSerGlnAlaVa 2332
Db 7062 GATCTTAATATGCTCCCGCATACCGGAACAGGCTTATGACTATAGAGCCAGCGGT 7121
QY 2332 lGlnGlYhiASerThrASnMetGlYleuGlYAlleIleArgYsAlaLeuMetGl 2352
Db 7122 GCAGGAACATGCCAGCAACCAATGGGGCTGGAGCCATTAATTAAGAAAGCACTCATGGG 7181
QY 2352 ylybTYrASnGlYnTgluJmSerProProleuSerAlaASnAlaPheAnProleuAs 2372
Db 7182 TAAATATGACCAAGGGAGAGTCCCGCGCTCAGCGCCCAATGCTTTAACTTCGAA 7241
QY 2372 naIaSerAlaSerleuProAlaIleMetProIleThrAlaIleASnGlyTargSerAPhI 2392
Db 7242 TGCCAGTCCAGGCTGCTGCTATGCTCATTAACGCTGCTGACGAGAGTGTGCA 7301
QY 2392 sThrLeuThrSerProGlYglYglYlybAlaYleValSerGlyTargProSerSerAr 2412
Db 7302 CACACTACCTCGCAGGTGGCGGGGAGGCCCAAGGTCTCTGGCAGACCCAGCACCGG 7361
QY 2412 glybAlaYleSerProAlaProGlYleuAlaSerGlyASnTargProProSerValSerSe 2432
Db 7362 AAAAGCCAAATCCCGCGCGGCTGCGCATCGGAGACCGGCACTCTGCTCTC 7421
QY 2432 rValhiASerGlYlybASnYsASnTargThrProleuThrASnTargAlaTargJmAs 2452
Db 7422 AGTGCACTCGAGAGAGACTGCAACCGCGAGCGCTCAACCGCGTGTGGAGGA 7481
QY 2452 pATgProSerSerAlaGlySerThrProPheProTYrASnProleuIleMetArgLeuGl 2472
Db 7482 CAGGCCCTGTCGCGAGGTTCCAGGCATTCCCTTAACCCCTGATCATGGCGTGA 7541
QY 2472 naIaGlyValMetAlaSerProProProProGlYleuProAlaGlySerGlyProleuAl 2492
Db 7542 GGGGGGTGTCATGCTTCCACCCCGCGGCTTCCCGCGGGGCGCGGCGCTCGC 7601
QY 2492 aglyProhiASnAlaThrASnGluJmProPlyProleuLeuCYSerGlnTYrGluTh 2512
Db 7602 TGGCCCCCAGCAGCGCTGGGACGAGGACCAAGCCACTGCTGTGAGTACGAGAC 7661
QY 2512 rLeuSerASnSerGlu 2517
Db 7662 ACTCTCGAGACGCGAG 7677

RESULT 9
US-10-723-860-6114/c
; Sequence 6114, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Gineburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193 .NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 6114
; LENGTH: 9079
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-6114
Alignment Scores:
Pred. No.: 0 Length: 9079
Score: 12845.50 Matches: 2473
Percent Similarity: 98.10% Conservative: 6
Best Local Similarity: 97.86% Mismatches: 20
Query Match: 97.20% Indels: 31
DB: 21 Gaps: 3
US-09-522-753-5 (1-2517) x US-10-723-860-6114 (1-9079)
QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTargAlaThrGluProArgTYrPro 20
Db 8552 ATGTGGGGCTCCACACAGCGCTGTGGCAGACGTGAGGGGCCATGAGCCCGCTTACCG 8493
QY 21 ProhiASerLeuSerTYrProValGlnIleAlaTargThrThiASnAPValGlyLeuLeu 40
Db 8492 CCCACAGCTTCTTCTACCAAGTGCAGATGCCCGGACGACACAGCGAGCTGGGCTCTG 8433
QY 41 GluTYrGlnhiASerASnTargTYrAlaSerThiASnSerProGlYSerTleIleGln 60
Db 8432 GAGTACCAAGACACACTCCCGCACTAGCTCCACCTGTGCGCGCTCATCATCAG 8373
QY 61 ProGlnTargTargTargProSerLeuLeuSerGluPheGlnProGlYASnGluTargSerGln 80
Db 8372 CCCAGGGGGAGGCGCTCTGCTGTCTGAGTTCCAGCCCGGGAATGAACGTTCCAG 8313
QY 81 GluLeuhiASnLeuTargProGlYSerThiASerTYrLeuProGlYleuGlyYsSerGluMet 100
Db 8312 GAGCTCCACCTGGGCGGACAGATCCCACTATCCTGCCAGCTGGGGAAGTCAGAAATG 8253
QY 101 GluPheIleGluSerTylASnTargTargLeuLeuLeuProleuProleuLeuTargPro 120
Db 8252 GAGTTATTGAAGCAAGCGCTCGCTGAGAGCTGCTGTAACCTTCAACCCCTGTGCGACG 8193
QY 121 SerProleuLeuAlaThrGlyGlnProAlaGlySerGlyASnProleuThrTylASnAPArgSer 140
Db 8192 TCACCTGCTGTGCGCACGGGCGAGCTGCGGATCTGAAGACTTCAACCAAGACCGGCC- 8134
QY 141 LeuThrGlyYleuGluProValSer-ProProSerProProhiASnThrASnProGlYle 160
Db 8133 CTGACGGGCAAGCTGGAACCGGTGTCTCCACCAACCCCGCACACTGACCTTGAGCT 8074
QY 160 uGluLeuValaProProArgLeuSerTylSerGlnJmleuIleGlnASnMetASnTargVala 180
Db 8073 GAGGTGTGTGCGCCAGGCTGTCCAAAGAGAGCTGATTCCAAAACCTGGAACCGCGTGA 8014
QY 180 pATgTylIleThrMetValGluGlnJmleuIleSerTylLeuYleYleGlnJmle 200
Db 8013 CCGAGAGATCAACATGTTAGAGCAGATCTTAAAGCTGAAGAAGACAGCAACGCT 7954
QY 200 uGluGluJmAlaIleYbProProGlYleuYbProValSerProProIleGln 220
Db 7953 GAGAGAGAGAGCTGCCAAGCGCGCCGAGCTGGAACCGGTGTCAACCGCGCCATCGA 7894
QY 220 uSerTylASnTargSerLeuValGlnIleIleTYrASnGluASnTargTylYbAlaGluAl 240
Db 7893 GTCGAAGACCGGAGCTGTGTGAGTCAATCTTACGAGAGAAACCGGAAGAGGCTGAAGC 7834
QY 240 AlaIleASnGlyleuGluGluGlyProGlnValGluLeuProleuTYrASnGlnPr 260

|||||
Db 7833 TCACATCCGATTCTGGAGGCTGGGGCCCGCAGGTGAGAGCTGCCCTGACCAACAGCC 7774
OY oSerAspThrArgIntYrHiEgLuAniLeuYsiLeaAnGlnAlaMetArgLysIle 280
Db 7773 CTCGACACCCGGCATTCATGAGAACATCAAAATTAACACAGGCATCGAGAAAGCT 7714
OY uileuYrPhelysArgArgAnhiAlaArgLysGlnTrpLysGlnLysPheCysG 300
Db 7713 AATCTTGATCTCAAGAGAGAAATCACCGCTCGAAACAATGGAGACAGAAAGTTCTGCCA 7654
OY 300 nArgTrpAspGlnLeuMetGluAlaLeuGluLysIleValGluArgTILEGluAsnAsnPr 320
Db 7653 GCGCTATGACCAAGCTCATGAGAGCCCTGGAGAAAGAGTGGAGCGCATGAGAAACACC 7594
OY 320 oArgArgArgAlaLysGlnSerLysValArgGluTrpLysGlnLysPheProGlu 340
Db 7593 CCGGCGCGGGCCAGAGAGCAAGAGTTCGCGAGTCTACAGAGACAGTCCCTAGAT 7534
OY 340 eArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgLysSerG 360
Db 7533 CCGCAAGACGCGGAGCTGAGAGAGCGCATGAG---AGGTTGGCCAGCGGGGAGTGG 7477
OY 360 yLeuSerMetSerAlaAlaArgSerGluHiSGluValSerGluIleIleAspGlyLeuSe 380
Db 7476 GCTGTCCATGTGCGCCCGCCGACGAGACGAGAGTGTCAAGATCATGATGCGCTTC 7417
OY 380 rGluGlnGlnLysLeuGluLysGlnLeuArgGlnLeuAlaValIleProProMetLeuTr 400
Db 7416 AGAGCAGAGAACCTGAGAGACAGATGCCCGACGTGGCCGTATCCCGCCCATGCTGTA 7357
OY 400 rAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMe 420
Db 7356 CGACCGTGAACGACGAGCATCACTTCATCAACATGAACGGGCTTATGGCCGACCCCAT 7297
OY 420 LysValIleTrpAspArgGlnValMetLeuMetTrpSerGluGlnGluLysGluThrPh 440
Db 7296 GAAGGTGTCAAAGACCGCAGGTTCATGAACATGTGAAGTGAACAGAGAAAGACCTT 7237
OY 440 eArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
Db 7236 CCGGAGAAAGTTCATGACGATCCCAAGAACTTTGGCTGATGCACTCATTTCTGGAGAG 7177
OY 460 gLysThrValAlaGluCysValLeuTrpTrpLysLeuThrLysLysAsnGluAsnTrpLys 480
Db 7176 GAAGACAGTGGCTGAGTGGCTCTCTATTACTGACTGACTAAGAGATGAGAATCATATA 7117
OY 480 sSerLeuValArgArgSerTrpArgArgArgLysLysSerGlnGlnGlnGlnGln 500
Db 7116 GAGCGCTGTGAGACGAGACTATCGGGCGCGCGCAAGAAC--CAGCAGCAAACAAAAG 7059
OY 500 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
Db 7058 CCGGACGACGACGCCACCAACGACGACGACGACGACGACGACGACGACGACGACGAC 6999
OY 520 sAspGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLys 540
Db 6998 AGATGAGAAAGAGAAAGAAAGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6939
OY 540 nAspLysGluLysPheLeuLysGluLysTrpAspAspThrSerGlyGluAspAsnAspG 560
Db 6938 CGACAAAGGAAGACTCTCTCAAGAGAAAGACGACGACCTCAGGGGAGAGACACACGA 6879
OY 560 uLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysG 580
Db 6878 GAAGGAGGCTGTGGCTCCAAAGCGGCAAACTGGCCACACGACGAGAAAGACCAAGG 6819
OY 580 yArgIleTrpArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnG 600
Db 6818 CCGCATCACCCGCTCATGTGCTATAGAGCCCAACAGAGAGAGAGCCATCACCCCCAGCA 6759
OY 600 nSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluLue 620

Db 6758 GAGCCCGACGCTGGCTCCATGAGAGTGAATGAGATTCTCCCTGGACAGAAAGAAAT 6699
OY 620 rGluThrAlaLysLysGlyLeuLeuGluHiSGluValArgAsnTrpSerAlaIleAlaArgMe 640
Db 6698 GGAAACAGCCAAAGAAAGTCTCTCGMAACAGCGCGCAACTGCTGGCCATCGCCCGAT 6639
OY 640 tValGlySerLysTrpThrValSerGlnCysLysAsnPheTrpPheAsnTrpLysLysArgG 660
Db 6638 GGTGGGCTCCAAAGACTGTGTCCAGTGAAGAACTTCACTTCACTCAACAGAAAGAGCA 6579
OY 660 nAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluLysArgAsnAlaAr 680
Db 6578 GAACCTCGATGAGATCTTCACAGACACAAAGCTGAAGATGAGAAAGAGAGAACCCGG 6519
OY 680 gArgLysLysLysValaProAlaAlaAsnSerGluGluAlaIlePheProValIle 700
Db 6518 GAGAAAGAAAGAAAGCGCCGGCGGCGCAGAGAGAGGTGCTTCGCCGCCGTGGT 6459
OY 700 lGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluG 720
Db 6458 GAGAGTGAAGATGAGAGCGTCGGCGCTGACGCGAAATGAGAGAGATGTTGAGGA 6399
OY 720 uAlaGluAlaLeuHisAlaSerGlyAsnGluValProArgLysGluCysSerGlyProAl 740
Db 6398 GCGTGA-----GCTGAA-----GC 6390
OY 740 aThrValaAsnAsnSerSerAspThrGlnSerIleProSerProHisTrpGluAlaAlaLys 760
Db 6389 CACTGTCAACAAACAGTCAAGACACCGAGAGCATCCCTCTCTCTCAACATGAGGCGCCCA 6330
OY 760 sAspThrGlyGlnAsnGlyProLysProProAlaThrLeuGluAlaAspGlyProProPr 780
Db 6329 GACACAGGAGAAAGGGCCCAAGCCCAACACCTGGGCGCGAGCGGCGACCCCC 6270
OY 780 oGlyProProTrpProPro-ArgArgTrpSerArgAlaProIleGluProTrpProAlas 800
Db 6269 AGGCGCACCCACCCACACCGAGAGACATCCCGGC-CCCACTAGTGTCAACCCCGGCT 6211
OY 800 eArgLysAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProPro 820
Db 6210 CTGAAGCCACCTTAGCCCTTAGCCCCCAACAGCACCCCAATTCCTCTTCACTTCCTC 6151
OY 820 roValValaProLysGluGluLysGluGluGluGluGluGluGluGluGluGluGluG 840
Db 6150 CTGTGTCTCCCAAGAGAGAGAGAGAGAGACCGCACAGCGCCCAAGTGGAGAGAG 6091
OY 840 lYglGluGluGlnLysProProAlaAlaGluGluLysLeuAlaValaAspThrGlyLysAlaG 860
Db 6090 GGGAGAGCAAGAGCCCCCGCGCTGAGAGAGTGGCACTGGACACAGGAAAGCCGAGG 6031
OY 860 lProValLysSerGluCysTrpGluGluAlaGluGluGlyProAlaLysGlyLysAspA 880
Db 6030 AGCCGCTCAAGAGCGAGTGCACGAGAGAAAGCGAGAGGGCGCGCCAAAGGCAAGAGCG 5971
OY 880 lGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyG 900
Db 5970 CCGAGGCGCTGTGGCGCACGGCCGAGAGGGCGCTCAAGGACAGAGAAAGAGAGGGCGGA 5911
OY 900 eArgLysArgLysThrAlaLysSerSerGlyValaProGlnAspSerAspSerSerAlaTr 920
Db 5910 GCGGAGGGCCACCAAGCCAAAGAGCTCGGGCGCCCCCAAGACAGCACTCAAGTGCCA 5851
OY 920 hCysSerAlaAspGluValaAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSer 940
Db 5850 CCTGCACTGACAGACGAGTGAAGAGCGAGGGCGGCAACAAAGACCGGCTGCTGCC 5791
OY 940 roArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysP 960
Db 5790 CAAGGCCAGCTCTCCACCCGACCTGGCGAGACCCCGGGCCCAATGTGCTCAACCCCAAG 5731
OY 960 roLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrL 980
Db 5730 CACTGGACTTGAAGCACTGAAGCAAGCGGCTCCATCCCCCCTCAGGTCAACA 5671

QY 980 yEValHISgluProProArgIuAspAlaAlaProThrLysProAlaProProAlaProP 1000
Db 5670 AAGTCCATGAGCCCCCGGGAGAGACGACGCTCCACCAAGCCAGCTCCCCAGCCCCAC 5611
QY 1000 roProProGlnAanLeuGlnProGlnSerAspAlaProGlnGlnProGlnSerSerProA 1020
Db 5610 CGGCACCGCAAAACCTGACACCGGAGAGACGCGCTCTGACAGAGCTGGAGAGAGCCCC 5551
QY 1020 rGgIlylSerSerArgSerProAlaProProAlaAspLysGlu----- 1033
Db 5550 GGGGCAAGACAGAGAGCCGCGACCCCGCCGCAAGAGAGGAGAGAGAGCCGTGTCTT 5491
QY 1034 -----AlaPheAlaAlaGluAlaGlnLysLeuProGlnYAspProProCysTrpThrSerg 1052
Db 5490 TCCAGAGCTTCGAGCGGAGGCGCCAGAAAGCTGCTGGGAGACCCCCCTGCTGGACTTCCG 5431
QY 1052 lYleuProPheProAlaProProArgIuValIleLysAlaSerProHISAlaProAsp 1072
Db 5430 GCCTGACCTTCCCGTGCCTCCCGGTGAGGTGATCAAGGCTCCCGCATGCCCCGAGC 5371
QY 1072 roSerAlaPheSerTrpAlaProProGlnYHISProLeuProLeuGlnLysHISAspThra 1092
Db 5370 CCTCAGAGCTTCTCTAGGCTTCACTGTGTCACCACTGCTGGGCTCCATGACACTG 5311
QY 1092 lAArgProValLeuProArgProProThrIleSerAsnProProProLeuIleSerSera 1112
Db 5310 CCGGCGCGCTTCCGCGCGCGCCACCCACATCTCCAAACCGGCTCCCTCATCTCTCTG 5251
QY 1112 lAlYHISProSerValLeuGlnArgGlnIleGlnAlaIleSerGlnGlnLysSerValG 1132
Db 5250 CCAAGACCCCGACGCTCTCGAAGGCAATAGGTGCATCTCCCAAGATGTCGATCC 5191
QY 1132 lIneuHISValProTrpTrpSerGlnHISAlaLysAlaProValGlnProValThreMetGln 1152
Db 5190 AGCTCAGAGCTTCTCTAGGCTTCAAGACATGCAAGGCGCGGTGGGCTCTGATGAGG 5131
QY 1152 euProLeuProMetAspProLysLysLeuAlaProPheSerGlnValLysGlnGlnL 1172
Db 5130 TGGCCCTGCCATGAGACCCCAAAAGCTGSCACCTTCAGCGAGTGAACAGAGAGCAGC 5071
QY 1172 euSerProArgIlyGlnAlaGlnIlyProProGlnSerLeuGlnValProThraAlaGlnL 1192
Db 5070 TGTCCCAAGCGGAGGAGCTGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 5011
QY 1192 lAserValLeuArgIlyThraAlaLeuGlnYserValProGlnGlnYserIleThraLysGln 1212
Db 5010 CGTCCGTGAGAGAGGAGAGCTGTGGCTCAGTTCGGGCGGAGAGCATCACAAAGGCA 4951
QY 1212 lAserSerThraArgValProSerAspSeraIleThraTrpArgIlySerIleThraHISg 1232
Db 4950 TTCCAGACACAGGAGTGCCTCGGACAGCGCATCATACCGGCTCATACCCACG 4891
QY 1232 lYThraProAlaAspValLeuTrpLysGlnYThraIleThraArgIleIleGlnLysAspSerp 1252
Db 4890 GCACGCGCAGGTGAGTCTGTACAGGAGCAGCATACAGATCATGGGAGAGAGAGC 4831
QY 1252 roSerTrpLeuAspAlaArgIlyArgIlyAspSerLeuProLysGlnYHISAlaIleTrpGln 1272
Db 4830 CAGGTGCTTGGACCGCGCGGAGAGAGCTGCGCCAGAGGCGCAGTCACTTACGAG 4771
QY 1272 lYl 1292
Db 4770 GCAGAAAGGCGCAGCTTGTCTTATGAGGAGTGCATGTCTGTGACCCAGTCTCCAGG 4711
QY 1292 lAspArgIlyArgSerSerSerGlnYProProHISGlnThraAlaAlaProLysArgThra 1312
Db 4710 AGGACGCGCAAGACACTCAGAGACCCCATGAGAGCGCGCCCAAGCGCACTTATG 4651
QY 1312 spMetSerGlnGlnYArgValGlnYArgAlaIleSerSerAlaSerIleGlnGlnLysLeuMetG 1332
Db 4650 ACATGATGAGGCGCGCTGGGAGAGAGCATCTCTCAGCAGCATGAGAGTCTCATGG 4591

QY 1332 lYArgAlaIleProProGlnUArgHISerProHISISLeuLysGlnGlnHISISleAla 1352
Db 4590 GCGGTGCATCCCGCGGAGCGAGACAGCGCCCAACCACTCTAAAGACAGACACATCC 4531
QY 1352 rGgIlySerIleThraGlnGlnIleProArgSerTrpValGlnAlaGlnGlnAspTrpLeuA 1372
Db 4530 GCGGCTCATACACAAAGGAGTCCCTGCTCTACGTGAGGAGCAGAGAGACATCACTGC 4471
QY 1372 rGArgGlnAlaLysLeuLeuLysArgGlnGlnIlyThraProProProProProSerArg 1392
Db 4470 GTGAGAGGCGCAAGCTCTTAAGCGGAGGAGCAGCTCCGCCCCACCGCTTCAGCGG 4411
QY 1392 spLeuThraGlnAlaTrpLysThraGlnAlaLeuGlnProLeuLysLeuLysProAlaHISg 1412
Db 4410 ACCTGACCGAGGCTTCAACAGCAGGCTGCGGCCCTCGAAGCTGAACCGGCGCATG 4351
QY 1412 lGlnLysLeuValAlaThraValLysGlnAlaGlnYArgSerIleHISGlnIleProArgGln 1432
Db 4350 AGGAGCTGTGGCGCACCGGTGAGAGGCGGCGCTCATCATGAGATCCCGCGCAGG 4291
QY 1432 lUleuArgHISThrProGlnUleuProLeuAlaProArgProLeuLysGlnGlnYserIleT 1452
Db 4290 AGCTGGGCGCACGCGCGAGCTCCCTGGCGCGCGCGCTCAAGAGAGGCTTCATCA 4231
QY 1452 hrGlnGlnYThrProLeuLysTrpAspThraGlnYlaserThraThraGlnYserLysHISAla 1472
Db 4230 CGAGAGGCGACCCCGCTCAAGTACGACACCGCGCTGACCACTGGCTCCAAAGAGACG 4171
QY 1472 spValArgSerLeuIleGlnYserProGlnYArgThraPheProProValHISProLeuAsp 1492
Db 4170 AGCTAGGCTCTCTCATGCGGAGCGCGCGCGGCGGCTCCACCGCGTGCATCGGATG 4111
QY 1492 alMetAlaAspAlaArgAlaLeuGlnUArgAlaCysTrpGlnGlnYserLeuLysSerArg 1512
Db 4110 TGAATGCCAGCGCGCGGAGCTGGAACGTGCTGTACGAGAGAGCTGAAGAGCGGCG 4051
QY 1512 roGlnThraLaserSerSerGlnYserIleAlaArgGlnYAlaProValIleValProG 1532
Db 4050 CAGGAGCGCGAGCACTCGGGGGGCTCATTCGCGCGCGCGCGCTCATGTGTGCTG 3991
QY 1532 lUleuGlnYlYsProArgGlnSerProLeuThraTrpGlnAspHISGlnYAlaProPheAlaG 1552
Db 3990 AGCTGGGTAAAGCGCGCGAGAGCGCCCTGACCTATGAGACACAGGAGCGCCCTTTC 3931
QY 1552 lYHISLeuProArgGlnYserProValThreTrpArgGlnUProThraProArgLeuGln 1572
Db 3930 GCCACTCCACAGAGTTCGCGGTACCATGCGGAGGCCAGCGCGCTGAGAGAG 3871
QY 1572 lYserLeuSerSerSerLysAlaSerGlnAspArgLysLeuThraSerThraProArgGln 1592
Db 3870 GCAGCTTTCGTCCAGAGCATCCAGAGCCGAAAGCTGAGTGCAGCGCTCGTAGA 3811
QY 1592 lAlaLysSerProHISerThraValProGlnHISISProHISProIleSerProTrpG 1612
Db 3810 TCGCAAGTCCCGGACACACACCGTGCAGAGCACCAACCCCATCTCCGCTATG 3751
QY 1612 lUHISLeuLeuArgGlnYlaserGlnValAspLeuTrpArgSerHISleProLeuAlaP 1632
Db 3750 AGCAGCTGCTCGGGCGTGAAGTGGGAGCTGTATGCAAGCCACATCCCGTGGCT 3691
QY 1632 heAspProThraSerIleProArgGlnIleProLeuAspAlaAlaAlaAlaTrpTrpLeuP 1652
Db 3690 TCGAGCCCACTCCATACCCCGGAGATCCCTGTGACCGAGCGCTGCTTACTCTGC 3631
QY 1652 roArgHISleuAlaProAsnProThraTrpProHISLeuTrpProProTrpLeuIleArgG 1672
Db 3630 CCGGACACCTGGCGCCCAACCCCACTTACCGGAGCTGTAACCACTCACTCATCGCG 3571
QY 1672 lYThraProAspThraAlaAlaLeuGlnUAsnArgGlnThraIleIleAsnAspTrpIleThs 1692
Db 3570 GCTACCCGACAGGCGGCGGTGAGAACCGGAGAGCATCATCAATGACTATCATCACT 3511
QY 1692 erGlnGlnMetHISISAsnThraAlaThraAlaMetAlaGlnArgAlaAspMetLeuArg 1712

[illegible]

Db	2430	GGGATGAGACCGCTCAGACCCCTGTGAGCTCAACCGAGTGCACCCAGCAACGAAGGGAGCTCCCA	2371
Qy	2072	YSHiSleuGlUGLUleuAaplyrSerSHiSleuGlUGLUleuAapProLyGlInP	2092
Db	2370	AGCACTTGAAAGAGCTCGCAAGAGCACTTGAGGGAGGAGCTGGCGGCCCAACAGCAG	2311
Qy	2092	LYProValLYSleuGlYGlYGlUaLaLaHiSleuProHiSleuAArgProLeuProGlus	2112
Db	2310	GGCCCGTGAAGTTGGCGGGGAGGCGGCCACTCCCAACACTCGGGCGCTGTGAGA	2251
Qy	2112	erGlInProSerSerSerProLeuLeuGlInThAlaProGlYValLYeGLYHiGLInArgV	2132
Db	2250	GCCACGCCCTCGTCACGCCGTGCTTCAACACCGCCACGAGGAGTCAAAAGGTACCAAGGGG	2191
Qy	2132	aIValThreunLaGlnHiSiLeSerGlYValILeThrgInaApTYrThArgHiSHiSP	2152
Db	2190	TGGTCACTTGCGCCAGCAATCAGTAGGTATATCACAGACTACACCGGGCACACACC	2131
Qy	2152	roGlInLeuSerAlaProLeuProAlaProLeuTYrSerPheProGlYAlaSerCysP	2172
Db	2130	CACAGAGCTCAGCGACCCCTGCGCCGCCCTCTACTCTTCCTGTGGGGCAGCTGCC	2071
Qy	2172	roValLeuAspLeuAArgAArgProProSerAspLeuTYrLeuProProAspHiGLY	2192
Db	2070	CGGCTCTGAACCTCGCGCGCCCAACCAAGTACTTACCTTCGCCGCCCGGACATGGTG	2011
Qy	2192	LaProAlaArgLYSerProHiSserGlUGLYLYuSArgSerProGlProAsnLYST	2212
Db	2010	CCCCGGCCGTGGCTCCCCCAGAGGAGGGGGCAAGAGTCTCCAGAGCCAAACAAGA	1951
Qy	2212	hSerValLeuGlYGlYGlUaApGLYLeuProValSerProProGlUGLYMet	2232
Db	1950	CGTCGGTCTTGGGTGGTGTGAGGAGCGATTGAACCTGTGTCCACCGAGGGCATGA	1891
Qy	2232	hGlUProGlYHiSserArgSerAlaValTYrProLeuLeuTYrArgAspGlYGlUGlnt	2252
Db	1890	CGGAGCCAGGGCACTCCCGAAGTCTGTGATCCCGCTGCTGTCACGGGATGGGGACGA	1831
Qy	2252	hGlUProSerArgMetGlYSerLYSerProGlYanThrSerGlnProProAlaPheP	2272
Db	1830	CGGAGCCAGAGGAATGGGCTCCAACTCTCCAGGCAACACACGAGCCGACGCTTCT	1771
Qy	2272	hSerLYSleuThrgIuSerAnSerAlaMerValLYSerLYSLGInLLeuSnt	2292
Db	1770	TGAGCAAGCTGACCGAGACAACTCCGCCATGTCTCAAGTCCAAAGACAAAGATCAACA	1711
Qy	2292	YALysLeuAnThrHiSAsnAArgAnGlUProGlUTYrAnSiLeSerGlnProGlYThrg	2312
Db	1710	AAAGCTGAACAACCCAAACCGGAATGAGCTGAATATACATATCAGCAGACTTGAGCAGG	1651
Qy	2312	IuLePheAnMetProAlaILeThnGLYThrgLYeUeWetThTYrArgSerGlnaLay	2332
Db	1650	AAATCTTCAATATGCCCCGCATCAACGGAAACAGGCTTTAAGCTTAAGAAGCCAGGCGG	1591
Qy	2332	aIGlNGlUHiSiLaSerThrAsnMetGlYLeuGlUaLaIeILeArgLYaLaLeuMetG	2352
Db	1590	TGCGAGAAATGCGAGCAACCAATATGGGCTGGAGGCCAATATTAGAAAGCACTATGG	1531
Qy	2352	LYLySTYrAspGlInTYrGlUGlUSeProProLeuSerAlaAnLaIaPheAnProLeuA	2372
Db	1530	GTAATATATAGCAGTGGGAAGAGTCCCGCGCTCAGCGCCAAATGCTTTTAAACCTCTGA	1471
Qy	2372	snAlaSerAlaSerLeuProAlaLaMerProILeThrLaLaIaAspGLYArgSerAspH	2392
Db	1470	ATGCCAGTCCAGCTCGCGCGCTGATGCGCAATACCGTGTCTGACGAGCAGAGTACCC	1411
Qy	2392	iSThrLeuTherSerProGlYGlYGLYLYuSAlaYSValSerGlYArgProSerSera	2412
Db	1410	ACACACTCACTGCGCAGAGTGGGGGGGAGGCAAGGTCTGTGCGACACCGCAGAGCC	1351
Qy	2412	rgLYaLaLYSerProAlaProGlYLeuAlaSerGLYAspAArgProProSerValSers	2432
Db	1350	GAAGAGCCAAAGTCCCGGCGCGGCTGTGCATCTGGAGCCGGGCAACCTCTGTCTCTCT	1291

460 ArgLysThrValAlaGluCysValLeuTyrTyrLeuThrLysAsnGluAsnTyr 479
Db AGAAAGACGGTGGCTGAGTGTCTCTCTATTACTGACCAAGAAAGATATAATTC 1437
Qy LysSerLeuValArgArgSerTyrArgArgGlyLysSerGlnGlnGlnGln 499
Db AAGACTTGTGTAGGGGAGGAGCTATCGGCGCCGTGGCAAGACGACGACGACGACG 1497
Qy GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 519
Db CAACAACAGACAGACGACGACGACGACGACGACGACGACGACGACGACGACG 1545
Qy LysAspGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLys 539
Db AAG 1605
Qy AsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyLysAspAsp 559
Db AACGAGAGAGAGAGAACTCAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1665
Qy GlnLysGluLysValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLys 579
Db GAGAAAGAGAGCGGTGGCTTCGCAAGGCGCGCAAACTGCAACGACCGACGCGCAAA 1725
Qy GlnArgLysThrArgSerMetAlaAsnGluAlaAsnSerGluLysAlaThrThrProGln 599
Db GCGCGATACCGCGCTCCATGCGCAACGAGCCACATGAGAGAGACACCGCCACAG 1785
Qy GlnSerAlaGluLeuAlaSerMetGluLeuAsnGlnLysSerArgTrpThrGlnGlnGln 619
Db CAAAGTTCAGACTGGCTTCCTCATGAGATGAACGAGAGTTCTGCTGGACTGAGAGAG 1845
Qy MetGluThrAlaLysLysGlyLeuLeuGlnLysGlyArgAsnTrpSerAlaAlaAlaArg 639
Db ATGGAGACGCAAGAGAGAGCGCTCTGGAACATGGAGAGAGATGGTCAAGCATGGCCGC 1905
Qy MetValGlySerLysThrAlaSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArg 659
Db ATGGTGGGCTCCAGAGCGGTGTCCAGGTATGAATCTTCACTTCAACTACAAAGAGAG 1965
Qy GlnAsnLeuAspGluLysLeuGlnGlnLysLysLeuLysMetGluLysGluLysArgAla 679
Db CAGAACTGAGAGAAATCTTCAGCAGCACAAGCTAAAGATGGAGAGAGAGAGAGAG 2025
Qy ArgArgLysLysLysLysLysAlaProAlaAlaLysSerGluLysAlaLysPheProVal 699
Db CGAG 2085
Qy ValGluAspGluGlnMetGluLysSerGlyValSerGlyValAsnGlnGlnGlnMetValGln 719
Db GCTGAGAGAGAGAGAGATGAGAGATCAGGCGCAAGTCCCAATGAGAGAGAGAGAG 2145
Qy GlnAlaGlnAlaLeuHisAlaSerGlyLysGlnGlnValProArg--GlyLysCysSerGly 738
Db GAGGCGAG 2205
Qy ProAlaThrValAsnAsnSerSerAspThrGlnSerIleProSerProHisThrGlnAla 758
Db CCAAGCTGCTGTCAACAGAGCTCTGATCTGAGAGAGTCCCATCCCGGCTTCAAGAGCC 2265
Qy AlaLysAspThrGlyGlnAsnGlyProLysProProAlaThr-----LeuGlyAlaAsp 776
Db ACGAGAGAGAGAGCT-----GGGCGTAAACCACTGAGCACTGAGCACTTGGCGGCTGCC 2316
Qy GlyProProGlyProProProThrProProArgArgThrSerArgAlaProIleGluPro 796
Db ACCAG 2370
Qy ThrProAlaSerGlnAlaThrGlyAlaProThrProProProAlaProProSerProSer 816
Db TCCCAAGTCCCTGATGCCAGTGGCCCAACCATCCCAAGAGCTTCC--CCATCACTGCC 2427

817 AlaProProProValValProLysGlnGlnLysGlnGlnGlnGlnGlnGlnGlnGln 836
Db GCACCCCGGCTACTGTGAGACAAGAGATGAACAAGAGCCCGGCTCCAGCTCCAGCAG 2428
Qy ValGlnGlnGlnGlnGlnGlnGlnLysProProAlaAlaGlnGlnLeuAlaValAspThrGly 856
Db ACAGAGAGAGTCCAAAGAGAGAGAGTGTAGAGCCGAGAGAG-----ATCATGTGGGA 2538
Qy LysAlaGlnGluPro-----ValLysSerGlyThr 867
Db AAGCCAGAGAGAGCCGAGAGCTCTGAGAGAGCCCGGAGAGGTGAAGGTGACCAAG 2598
Qy GlnGlnAlaGlnGlnGlyProAla--LysGlyLysAspAlaGlnAlaAlaGlnAlaThr 886
Db GAGGAGACCGAGAGAGAGAGCTGAGACAAGCCCAAGGCGCACAGAGCCCTTAAACTGTG 2658
Qy AlaGlnGlyAlaLeuLysAlaGlnLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 905
Db TCTGAGGACACACTTAAGGTGAG-----GAGGCTGTAGCAAGGCAAGCTGTGACCAAG 2712
Qy AlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAspGln 925
Db GGTTCAGCTCAGTGTGCCACCGAGAGAGTGTCCAGTGTGCCACCTGACGTGCCATGAG 2772
Qy ValAspGlnAlaGlnGlyGlyLysLysAsnArgLeuLeuSerProArgProSerLeuLeu 945
Db GTGACGAAACCGGAGAGAGGTGACAAAGGCGAGGCTGTCTGTCAACAAAGCCGACCTCTC 2832
Qy ThrProThrGlyLysProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGln 965
Db ACCCGGCTGGAGATCCCGGCGCAGTACCTGCGCCAGAAACCGCTTGACCTTGAACAG 2892
Qy LeuLysGlnArgAlaAlaAlaThrProProIleGlnValThrLysValHisGluProPro 985
Db CTGAAGACGAGAGAGAGCGCGCATCCCGCTATC---GTCAACAGGTTCATGAGCCCGCC 2949
Qy ArgGluAspAlaAlaProThrLysProAlaProProAlaProProProProGlnAsnLeu 1005
Db CGGAGAGAGACAGATCCCGCAAGCGAGTTCCTCCGTGCGCTCCACCGCAGAGACCTTA 3009
Qy GlnProGlnSerAspAlaProGlnGlnProGlnLysSerSerProArgGlyLysSerArgSer 1025
Db CAGCAGAGAGGTGAGTGTCTCAGAGTGTGAGAGAGAGTCCAGTGGCAAGTCCGAGAC 3069
Qy ProAlaProProAlaAspLysGlu-----AlaPheAlaAla 1037
Db CCAAGTCCCTCGCGAG 3129
Qy GlnAlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProVal 1057
Db GAGGCGCA--AGCTACCGAGCTAGGCCCGCCAGCTGTCTATCGGCT--GCTTCCCAT-- 3183
Qy ProProArgGlnValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyr 1077
Db CTTCAACGGAGAGGTATCAAGACTTCACAGCGGCT--GACCT--CTCTTCTCTAC 3237
Qy AlaProProGlyHisProLeuProLeuGlnLysHisAspThrAlaArgProValLeuPro 1097
Db ACACCCCGGCTACACCGGTGCTCTGGGCTCCAGAGATAGCCCGGCGCGTCCGCA 3297
Qy ArgProProThrIleSerAsnProProProLeuLysSerSerAlaLysHisProSerVal 1117
Db CGTCCCGCC--ATCTTAACCCCGCCACCTTCATCTCTGCGCAAGATCCCGCGTA 3354
Qy LeuGlnArgGlnIleGlyAlaIleSerGlnGlnMetSerValGlnLeuHisValProTyr 1137
Db CTTGAGAGGACACTGGGTCCATCTCCAGGGAGTCAAGTCAAGTCTGTGTGCTTCCAC 3414
Qy SerGlnHisAlaLysAlaProValGlyProValThrMetGlyLeuLeuProLeuPheMetAsp 1157
Db TCGAGACATGCCAAG--CCCATGGGCTCTCAGCATGAGATGCGCCCTTGGCGTGAGAC 3471
Qy ProLysLysLeuAlaProPheSerGlyValLysGlnGlnGlnLysSerProArgGlyGln 1177

Db	3472	CTTAAGAGCTG-----	3483
Oy	1178	AlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGly	1197
Db	3484	-----GGG	3486
Oy	1198	ThrAlaLeuGlySerValProGlyGlySerIleThrIleGlyIleProSerThrArgVal	1217
Db	3487	ACAGCACT-GGCTCCGCCACCAAGTGAAGACATCACCAAGGGCGCT-CCCAAGTACCGGGCT	3544
Oy	1218	ProSerAapSerAlaIleThrTrpArgGlySerIleThrHisGlyThrProAlaAspVal	1237
Db	3545	GCAGACGGCCCC-----AGCTACAGAGCTCTTATCACCAAGC-ACGCCGAGACGCTC	3597
Oy	1238	LeuTrpIleGlyThrIleThrArgIleIleGlyGluAapSerProSerArgLeuAspArg	1257
Db	3598	CTCTACAAAGGATACATCAGCAGATCTCGTGGTAGGACAGCCCAAGTCGCTTGACCGG	3657
Oy	1258	GlyArgGluAapSerLeuProGlyHisValIleIleTrpGlyIleValGlyHisVal	1277
Db	3658	GCACAGAGAGACACCCCTGCCACAGGGCATGTCATCTAGAGGGCAAGAAAGCCACGTC	3717
Oy	1278	LeuSerTrpGlnGlyIleMetSerValThrGlnCysSerTrpGluAapGlyArgSerSer	1297
Db	3718	CTATCTCTTGAAGGTGATGTCCTGTGCACAGTCTTAAGGAGATGAAGAGGACGACG	3777
Oy	1298	SerGlyProProHisGluThrAlaAlaProLysArgThrTrpAapMetMetGluGlyArg	1317
Db	3778	TGGGCCCCACCCCATGATGATCGGCCCTTAACGACCTTATGACATATGAGAGCGCT	3837
Oy	1318	ValGlyArgAlaIleSerSerAlaSerIleGlyIleuMetGlyArgAlaIleProPro	1337
Db	3838	GTAGGACAGCATGCACTCATCGACCATAGAGGACATCATGGGCGCGGCATC---CCT	3894
Oy	1338	GluArgHisSerProHisIleLeuIleGluGlnHisHisIleArgGlySerIleThrGln	1357
Db	3895	GAGCGACACAGCCCC---CACCTCAAGAGAGCATCATCCACAGGCTTCATCAGCGAA	3951
Oy	1358	GlyIleProArgSerTrpValGluAlaGlnGluAapTrpLeuAspArgGluAlaIleLeu	1377
Db	3952	GGCATCCGAGAGTCTATGTGAGGCGGAGAGGATCTTACGGCGGGAAGCCAAAGCTC	4011
Oy	1378	LeuIleAspArgGluGlyThrProProProProProSerAspAapLeuThrGluAlaTrp	1397
Db	4012	TTGAAGCAGAAAGGAGACCAACTCCCCACACCACTCGGAGACTGATCGAAGACCTAC	4071
Oy	1398	LysThrGln-----AlaLeuGlyProLeuIleuIleuIleProAlaHisIleGlyLeu	1414
Db	4072	AAGGCCCGGCCCTCGACCTCTGGGTGCTCCCTGAAGCGAAGCCGACATCAGAGGGTGTG	4131
Oy	1415	ValAlaThrValIleValGluAlaGlyArgSerIleHisGluIleProArgGlnLeuArg	1434
Db	4132	GTAGCAATGTGAAGAGAGCGGCGGCTGTATCCATGAGATCCGAGAGAGACCTCGC	4191
Oy	1435	HisThrProGluLeuProLeuAlaProArgProLeuIleGlySerIleThrGlnGly	1454
Db	4192	CGCACCTGATGCTACCCCTGGCACACAGGCTCTGAAGAGAGGTTCCATACCCAGGGC	4251
Oy	1455	ThrProLeuIleTrpAspThrGlyAlaSerThrThrGlySerTrpHisAapValArg	1474
Db	4252	ACCCCACTCAAGTACGACTCTGGGGCACCTCCCACTGGCACCAAGAAACAGAGTGGC	4311
Oy	1475	SerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAapValMetAla	1494
Db	4312	TCATTCATCGGACAGCCCGGCGGCTTTCCCTCGCCCGACCCGCTGACATATAGCT	4371
Oy	1495	AspAlaArgAlaLeuGluArgAlaCysTrpGluGluSerLeuIleYSerArgProGlyThr	1514
Db	4372	GACGCGCGGAGCTGAGCGTGGCTCGTATAAGAGTCTGAAGAGCCGCTCAGGGAC	4431
Oy	1515	AlaSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGly	1534

Db	4432	AGCACTGATGACGGGGAGCTCCATACACAGTGGGGCTCCAGTCTGTCGCTGAACCTGGAC	4431
OY	1535	LYSPROATGINSERPROLEUTHThYrGLuaPhisGLyAlaProPhaAlaGLYHISleu	1554
Db	4492	AAGCCAGGCAAGGCCCATCTGACTTTAGAAACACACGGGGACCTTACCAAGTCACTTG	4551
OY	1555	PROATGlySERPROVALThMetArgGLuProThrPROATGleuGLuGLYSerleu	1574
Db	4552	CCAGGTGGCTCCCTGTGACACGAGGAGGCCACGCCACCCCTTCACGAAGGACGCTTC	4611
OY	1575	SERSErlySAIAsERGLuaSAPArguYsleuThrSERThrPROArgGLuIleAlaYs	1599
Db	4612	CTATCCAGCAAGGGGTCCACGAGCCGGAAGGTGACATTTACACCCCGGGAATGCCAAAG	4677
OY	1595	SERPROHISserThrValPROGLuHISHisPROHISProLIESerProTYrGLuHISleu	1614
Db	4672	TCCCAACACACACTGTGTCCCGAGACACACACCTCACCCCATCTCCCTTAGAGACACTTG	4731
OY	1615	LeuArgGLyAlaSERGLyValAspLeuTYrArgSerHISIlePROleuAlaPhaAspPro	1634
Db	4732	CTCCGGGGCGTGACTGTGTGATCTGATACCGTGTGATCATCCCATTTGGCTTTGACCCC	4791
OY	1635	ThrsertIlePROArgGLYIlePROleuAsp---AlaAlaAlaAlaTYrTYrLeuPROArg	1653
Db	4792	ACCTCCATACCCCGAGAGATCTCTCTGTGAGACACACAGCCGAGGCTACTACTCTCCCGG	4851
OY	1654	HISleuAlaPROAsnPROThrTYrPROHISleuTYrPROTYrLeuIleArgGLYTYr	1673
Db	4852	CACTTGGGCCCCAAGCCCCACCTTACCACCTGTAGCCACTTACCTCATCTGCGGGCTAC	4911
OY	1674	PROAspThrAlaAlaLeuGLuAsnArgGLnThrIleIleAsnAspTYrIleThrSERGLn	1693
Db	4912	CCTGACACGGGGCGCTGTGAGAACCGGCACAGCATCATCATGACTCATCACTCTCGCAG	4971
OY	1694	GLInethrSHISAsnThrAlaThrAlaMetAlaGLnArgAlaAspMetIleuArgGLYleu	1733
Db	4972	CAGATGACCAACCAACGGCTGCTCCCGCCATGAGCCCGAGGTGTGACATGTAGAGGAGTCTG	5031
OY	1714	SERPROArgGLuSERSErleuAlaLeuAsnTYrAlaAGlyPROArgGLYIleIleAsp	1733
Db	5032	TCAACCGGAGAGTCTGTGGTGGCCCTTAATTATTCGGTGGCCCAAGAGCATTTATGCAC	5091
OY	1734	LeuSERGLnValPROHISleuPROValIleuValPROProThrPROGLYThrPROAlaThr	1753
Db	5092	CTGTCCCAAGGCGACACCTGCGCTGTGGTGGCCACCAAGCAGGACACCCCTGCACAC	5151
OY	1754	AlaMetAspArgGLuAlaTYrLeuPROThrAlaPROGLnPROPhesERSErArgHISer	1773
Db	5152	GCCATTCACCGCCCTTGGCTACTCTCCCACTGCGGCCCCACCTTCACACGCGGCACAGT	5211
OY	1774	SERSErPROleuSERPROGLYGLYPROThrHISleuThrLYSPROThrThrThrSERSer	1793
Db	5212	AGCTACCGGCTGTCCCAAGAGGCGCCCACTCACTTAACCACTGCCACATCTTCA	5271
OY	1794	SERGLuArgGLuArgAspArgAspArgGLuArgAspArgGLuArgGLuLYSer	1813
Db	5272	TCCGAGGGGAGCGGAGAACTGAGCGGGAACGAGAC-----AAGTCC	5313
OY	1814	IleLeuThrSERThrThrThrValGLuHISAlaProIleThrPArgPROGLYThrGLuGLn	1833
Db	5314	ATCTCTCAAGCTTACCATCAAGTGGACATGCAACCTCTGGAGAACTGTGTACCGAACAG	5373
OY	1834	SERSErGLYserSERGLYserSERGLYGLYGLYGLYserSERSErArgPROAlaSer	1853
Db	5374	AGCAGCGGGGCT-----GGGGGACAGACCGGCCGCTCC	5409
OY	1854	HISerHISAlaHISGLnHISERPROIleSERPROArgThrGLnAspAlaLeuGLnGLn	1873
Db	5410	CACAGC-----CACAGACACTGGCCATCTCCCGGAGCCGAGACGCTTGCACAG	5463
OY	1874	ArgPROserValIleuHISAsnThrGLYMetGLYGLYIleIleThrAlaValAluPROser	1893
Db	5464	AGGCCACAGTGTGTGCAACAAGAGCATAAAGGCGCTGTATCACTCCGTGGAACCCGGC	5523

QY 1894 LysProThrValLeuArgSerThrSerThrSerSerProValArgProAlaIleAlaThrPhe 1913
DB 5524 AGCCCAACGCTCGTAGGTCACACTCCACTTCCGCTGTCGCCAGCTGCACATTC 5583
QY 1914 ProProAlaThrHisGlySerProLeuGlyGlyThrLeuAspGlyValAlaYrrProThrLeuMet 1933
DB 5584 CCACCTGCACACCCACGACCTTGGTGGCACTTGAAGGGGTTCACCTTACCTCATTG 5643
QY 1934 GluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProAlaGly 1953
DB 5644 GAGCCCGTCTCTGTACCCAGAGAGACTCTCGGGTGGCCCGGCCCAAGCGGCCCTGTG 5703
QY 1954 AspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSer 1973
DB 5704 GACGGGGCCATGCTTCCCTCACCAAAACCCCGGCCG-----GAGCCCGCTCC 5754
QY 1974 SerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThr 1993
DB 5755 TCACCCAGAGAGGCTCCAGGCCCATCTTGAACCCCCCACTCCAGCCACAGACC 5814
QY 1994 IleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProProAla 2013
DB 5815 ATGCGCCGACCCCAAGAAAGACCTTGCACCCCACTGCTCCAGTCCGAGCCGCGGG 5874
QY 2014 ProProAlaSerAlaSerAspProHisArgGlyLysThrGlnSerLysProPheSerIle 2033
DB 5875 ---CCCACTCGGCTCAGATCGACACGAAAGAAAGACTCAAGTAAACCTTTTCCATC 5931
QY 2034 GlnGluLeuGluLeuArgSerLeuGlyYrrHis---GlySerSerYrrSerProGluGly 2052
DB 5932 CAGGAATTGGAATTCGGTCTCTGGGGTTACCAAGTGAAGCTGGCAGCAAGCCCGATGG 5991
QY 2053 ValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLys 2072
DB 5992 GTGGAGCCCATCAGCCCGGTGAGCTCCCGACCTGACCCACAGCAAGGGGCTCTCAA 6051
QY 2073 HisLeuGluGluLeuAspLysSerHisLeuGluGlyLeuLeuArgProLysGlnProGly 2092
DB 6052 CCTCTGGAAGAGCTTAGAAGAGCACTTGGAAAGGGAAGCTGGGCAAGACGACGAGC 6111
QY 2093 ProValLysLeuGlyGlyGlyAlaHisLeuProHisLeuArgProLeuProGluSer 2112
DB 6112 CCATTAACCTCAGCGCGAGGCTGCCATCTCCACATCTGGGGCACCTGCCGAGAGC 6171
QY 2113 GlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgVal 2132
DB 6172 CAGCCCTCATCCAGCCCACTCTCCAGACTGCCCCAGGCATCAAAAGTCAACGAGGGTG 6231
QY 2133 ValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspYrrThrArgHisAsp 2152
DB 6232 GTCAACCTGGCTCAGACATCAGCAGAGTCAATACAGAGCAACACCGGACACCG 6291
QY 2153 GlnGlnLeuSerAlaProLeuProAlaProLeuYrrSerPheProGlyAlaSerGlyPro 2172
DB 6292 CAGCACTCAGTGGCCCTTCCCGCCCTCTTACTCTTTCCTCCGAGCACTCCCT 6351
QY 2173 ValLeuAspLeuArgArgProProSerAspLeuYrrLeuProProAspHisGlyAla 2192
DB 6352 GTGCTGGATCTTGCGGCCCAACCACTGACTTACTCTCCACCCCAACCACTGGCAC 6411
QY 2193 ProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnLysThr 2212
DB 6412 CCAGCCCGGGGATCCCCCACTGAAGGGGGCAAAAGTCCCAAGACCCAGCAAAACA 6471
QY 2213 SerValLeuGlyGlyGlyLysAspGlyIleGluProValSerProProGluGlyMetThr 2232
DB 6472 TCGGTCTCGGCGCAGTGGATGCCATTGACCTGTGTGCCACAGGGGCAATGAACT 6531
QY 2233 GluProGlyHisSerArgSerAlaValYrrProLeuLeuYrrArgAspGlyGluGlnThr 2252
DB 6532 GAGCCAGACATCTCGAGCGCTGTGTACCACTGCTGTATCGAAGCGGGAAACGAGG 6591

QY 2253 GluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePhe 2272
DB 6592 GAGCCC---AGGATGGGCTCTTAAGTCTCCAGGCAACACAGAGCGGCCACTTCTTC 6648
QY 2273 SerLysLeuThrGlnSerAsnSerAlaMetValLysSerLysGlnGlnIleAsnLys 2292
DB 6649 AGTAAGCTACTGAGAGCACTCCGCATGTGTGMACTGMAAGACAGAGATCAACAG 6708
QY 2293 LysLeuAsnThrHisAsnArgAsnGluProGluYrrAsnIleSerGlnProGlyThrGlu 2312
DB 6709 AAATCTAACACCAACCAACCGGAACGAGCAGAAATCAATATTGGCCAGCTGGAGAGAA 6768
QY 2313 IlePheAsnMetProAlaIleThrGlyYrrHisGlyLeuMetThrYrrArgSerGlnAlaVal 2332
DB 6769 ATCTTCACATCCCGCCCATCACTGAGCAGGCTTATACCTGTGAAGCCAGCGGTG 6828
QY 2333 GlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgGlyAlaLeuMetGly 2352
DB 6829 CAGAAACACGCGACACCAACATGGGGCTTGAAGCCATTATTAAGAAAGCATTGAGT 6888
QY 2353 LysYrrAspGlnTrrProGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsn 2372
DB 6889 AATATGATCAGTGGGAAGACCCCGCGCTCGGCGCAATGCTTTTAACCTCGAAT 6948
QY 2373 AlaSerAlaSerLeuPro---AlaAlaMetProIleThrAlaAlaAspGlyArgSerAsp 2391
DB 6949 GCAGGCGCAGTGTGCGCTGCTGTATGCCATTAACCACTGTGACGAGCGAGTGGAC 7008
QY 2392 HisThrLeuThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyYrrArgProSerSer 2411
DB 7009 CAGCACTCACTCCGCGAGTGAAGGTGGAAAGCCAAAGTCTTGCGAGACTTACAGAC 7068
QY 2412 ArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSer 2431
DB 7069 CCAAAAGCCCAAGTCGACAGCACAGGCTAAGGTCCGAGAGACGACCCCTTCTCTCC 7128
QY 2432 SerValHisSerGluGlyAspGlyAsnArgGlyThrProLeuThrAsnArgValTrrProGlu 2451
DB 7129 TCAGTACACTCAGAGGGGAGCTGCAATCCCGAAACCACTCAACCAAGTGTGTGGAG 7188
QY 2452 AspArgProSerSerAlaGlySerThrProPheProYrrAsnProLeuIleMetArgLeu 2471
DB 7189 GACGGCCCTCATCTGCAAGGTTCACGCCAATTCCTTCAACCTTTGATTAGAGCTA 7248
QY 2472 GlnAlaGlyValMetLysSerProProProProGlyLeuProAlaGlySerGlyProLeu 2491
DB 7249 CAGGACAGTGTATGGCTCCCGGCCCACTGGCTTGGCGAGGAGCGGCCCCCTA 7308
QY 2492 AlaGlyProHisHisAlaTrrAspGlyGluProLysProLeuLeuLysSerGlnYrrGlu 2511
DB 7309 GCTGTGCCACCAACCCCTGGGATGAGAGCCCAAGCACTGCTGTTCACAGTATGAG 7368
QY 2512 ThrLeuSerAspSerGlu 2517
DB 7369 ACACTCTCGAGACGCGAG 7386

RESULT 11
US-09-819-104A-4
; Sequence 4, Application US/09819104A
; Publication No. US20030027137A1
; GENERAL INFORMATION:
; APPLICANT: Chen, J. Don
; TITLE OF INVENTION: NOVEL NUCLEAR RECEPTOR COMPRESSOR MOLECULES
; TITLE OF INVENTION: AND USES THEREFOR
; FILE REFERENCE: UMG-030
; CURRENT APPLICATION NUMBER: US/09/819,104A
; PRIORITY FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/193,138
; PRIOR FILING DATE: 2000-03-29
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 8544

```

; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (160)..(7545)
; US-09-819-104A-4

```

Alignment Scores:

```

Pred. No.: 0 Length: 8544
Score: 10832.50 Matches: 2149
Percent Similarity: 87.94% Conservative: 90
Best Local Similarity: 84.41% Mismatches: 197
Query Match: 81.97% Indels: 117
DB: 10 Gaps: 33

```

US-09-522-753-5 (1-2517) x US-09-819-104A-4 (1-8544)

```

QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTyrAlaThrGluProArgTyrPro 20
DB 160 ATGTCAAGATTCACACAGCCTGTGGCACAGACATGGCGGCTGCTGAGCCCGCTACCCA 219
QY 21 ProHisSerLeuSerTyrProValGlnIleAlaGlnThrHisThrAspValGlyLeuLeu 40
DB 220 CCCCATGGCATCTCTCTACCCGGTGCAGATAGCCGCTCCACACAGACGTGGGGCTCTT 279
QY 41 GlnTyrGlnHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
DB 280 GAGTACCAACACACACCCCGGACTACACCTGTCACCCGCTTCATCATCCAG 339
QY 61 ProGlnArgArgArgProSerLeuLeuSerGlnIleGlnProGlyAsnGlnArgSerGln 80
DB 340 CCACAGAGGAGCGCGCTCTCACTGTCAGAGTTCCAGCTGGAGTGAACGGTCTCAG 399
QY 81 GlnLeuHisLeuArgProGlnLeuSerHisSerTyrLeuProGlnIleGlnTyrSerGlnMet 100
DB 400 GAGCTTCACCTGCGCCCTGAGTCCGACAGTTCTGCTGAGCGGCAAGCCCGACATA 459
QY 101 GlnPheIleGlnSerLeuArgProArgLeuGlnIleLeuValProAspProLeuLeuArgPro 120
DB 460 GAATTCACCGAGACAGACGCCCGCCCTGAGACTACTACCCGATACCTGCTGCGCCA 519
QY 121 SerProLeuLeuAlaThrGlnIleProAlaGlySerGlnAspLeuThrIleAspArgSer 140
DB 520 TCACCCCTGCTGGCCACTGGGACCGAGTGGTCTGAAGACCTTACCAAGACCGTAGC 579
QY 141 LeuThrGlyIleLeuGlnIleProValSerProProSerProProHisThrAspProGlnLeu 160
DB 580 CTGGCAGGCAAGCTGAGCCTGTGTCACTCCAGTCCCGCAGCGCTGAGCTGAGCTA 639
QY 161 GlnLeuValProProArgLeuSerLeuGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 179
DB 640 GAGCTGGCGGCATCTCTCACTCTCCAGAGAGAGCTGTATCCAGAACATTTGAGACCGGCTG 699
QY 180 AspArgGlnIleThrMetValGlnGlnGlnIleSerLeuLeuValGlyValGlnGlnIle 199
DB 700 GACCGTAGATTCACCATGTAGACACAGATCTCCAAGCTGAAGAGAGACAGCAACAG 759
QY 200 LeuGlnGlnGlnIleAlaIleValSerProProGlnIleProGlnIleSerProProIle 219
DB 760 TTGGAGAGGAGGCGCGCCAGCGCCGGAACCCGAGAGCGCTGTGTGCCACCAACCCATA 819
QY 220 GlnSerLeuHisArgSerLeuValGlnIleIleTyrAspGlnAsnArgLeuValGlnIle 239
DB 820 GAATTCAAAGACCGAGCGCTGTCTCAATCATCTACATGAGAACCGGAAGAACCGGAA 879
QY 240 AlaAlaHisArgIleLeuGlnGlnIleGlnIleProGlnIleGlnIleProLeuTyrAsnGln 259
DB 880 GCGGCAACACCGGATCTTAGAAGGCGCTGGGGCCCAAGGTGAGCTCTGTACCAACAG 939
QY 260 ProSerAspThrArgGlnIleTyrHisGlnIleValIleValIleAsnGlnIleMetArgLeuVal 279
DB 940 CCGTCTACACACCGCAGTACCATGAAACATCAAAATAAACACGAGCGATGCGGAGAGAG 999

```

```

QY 280 LeuIleLeuTyrPheLeuValArgAsnHisAlaArgLeuGlnIleTyrGlnIleValPheCys 299
DB 1000 CTGATCTTTGACTTTAAGCCGAGGACACACCGCGCAAGACGTGGAAACAGCGCTTCTGC 1059
QY 300 GlnArgTyrAspGlnLeuMetGlnIleLeuGlnIleValValGlnIleGlnIleGlnIle 319
DB 1060 CAGCGCTATGACCGACTCATGTAGAGCGCTGGAGAGAAAGATGAGCCATGAGAAACAT 1119
QY 320 ProArgArgArgAlaValArgGlnSerLeuValArgGlnTyrTyrGlnIleValPheProGln 339
DB 1120 CCGGAAAGAGGCGCCAGAGAGACAGATGAGAGTGAAGTCTACGAAACAGTTCGCCGAG 1179
QY 340 IleArgGlnIleArgGlnLeuGlnIleValArgMetGlnSerArgValGlnIleArgGlySer 359
DB 1180 ATCCGACAGACGCGGAGCTGACGAGCGCATGACAGACAGCGGTGGCCAGCTGGCAGT 1239
QY 360 GlnLeuSerMetSerAlaAlaArgSerGlnHisGlnIleIleSerGlnIleIleAspGlyLeu 379
DB 1240 GGGCTCTCCATGTCTGGCTGCCCGCAGTGAAGCATGAGGTTTCTGAGATCATTTGATGGCTTG 1299
QY 380 SerGlnGlnIleValAsnLeuGlnIleValGlnIleMetArgGlnLeuValIleProProMetLeu 399
DB 1300 TCTGAGCAGAGAACTTGAAGACAGATGCGCACGCTGGCCGTGATCC--GCCATTTTG 1357
QY 400 TyrAspAlaAspGlnIleArgIleValPheIleAsnMetAsnGlnLeuMetAlaAspPro 419
DB 1358 TAGAGCCG -GACACAGACAGAGATCAATGTTATCATCACTGATGAGACTCATGATGAGCCC 1416
QY 420 MetIleValIleTyrIleAspArgGlnIleValMetAsnMetTyrSerGlnGlnIleValThr 439
DB 1417 ATGAGAGTCTTAAAGACCCCTGAGTTTACCAATGTGAGACGAGCAGAGAGGAGACCC 1476
QY 440 PheArgGlnIleValPheMetGlnHisProValAsnPheGlnIleIleIleAspPheLeuGln 459
DB 1477 TTTCCGTGAGAAATTATGACGACCCCTAAGAACTTTGGCCCTGATGCTCATTTCTTGAG 1536
QY 460 ArgIleThrValAlaGlnCysValLeuTyrTyrTyrLeuThrIleValAsnGlnIleAsnTyr 479
DB 1537 AGAAAGCGCTCCCTGATGTGTCTCTATTACTGACCAAGAAATGAAATTAATTC 1596
QY 480 LysSerLeuValIleArgArgSerTyrArgArgArgGlyLysSerGlnGlnIleGlnIle 499
DB 1597 AAGAGCTGTGTGAGCGGAGCTTATCGCGCGTGGCAAGACAGCAGCAGCAGCAG 1656
QY 500 GlnGlnGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 519
DB 1657 CAACACAGCAGACGACGACGAG-----ATGACAGGAGACAGCCAGAGAGAG 1704
QY 520 LysAspGlnIleValGlnIleValGlnIleValGlnIleValGlnIleValGlnIleValGlnIle 539
DB 1705 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1764
QY 540 AsnAspIleGlnIleValLeuLeuValGlnIleValTyrThrAspAspThrSerGlyIleAspAsnAsp 559
DB 1765 AACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1824
QY 560 GlnIleGlnIleValIleValIleSerIleValArgIleValThrAlaAsnSerGlnIleValArgIleVal 579
DB 1825 GAGAAAGAGGCGCTGTGCTTCAAGAGCGCGCAAAATGTCACCAAGCCAGGCGCGCAAA 1884
QY 580 GlnArgIleThrArgSerMetAlaAsnGlnIleValAsnSerGlnIleValIleThrProGln 599
DB 1885 GGCCTGTATCCGCGCTTCATGGCCAAAGAGCCAAACATGAGAGAGACAGCACCCACAG 1944
QY 600 GlnSerAlaGlnIleValAsnMetGlnLeuAsnGlnIleSerArgTyrThrGlnIleGlnIle 619
DB 1945 CAAGTTTCAGAGCTGCTTCATGAGAGATGAAGAGAGATTTCTGCTGAGATGAGAGAG 2004
QY 620 MetGlnThrAlaValValGlnIleValGlnIleValGlnIleValGlnIleValGlnIleVal 639
DB 2005 ATGAGAGACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2064
QY 640 MetValGlnIleSerIleThrValSerGlnCysValAsnPheTyrPheAsnTyrIleValArg 659

```

Db 2065 ATGGTGGGCTCCAGACCGCTGTCCAGTGTAAAGACTTCTACTTCAACTACAGAGAGAG 2124
 Qy 660 GlnaenleuAspGluIleuenglnIhIstlyleuIyMeGluIyysgluaraganaIa 679
 Db 2125 CAGAACCTGGAGAAATCTTACAGACCAACTTAAAGTGAAGAAAGAGAAAGCGCT 2184
 Qy 680 ArgatgIylylelylelysaIaProaIaIaIaIaSerGluIuIaIaIaPheProProIa 699
 Db 2185 CCGAGAGAAAGAAAGAACCCAGCTGCGGAGAGAGAGACACCTTCCACCTGCGC 2244
 Qy 700 ValIgluaSerGluIuMetGluIaSerGlyValaSerGlyAsnGluIuIuMetValIglu 719
 Db 2245 GCTGAGAGAAAGATGAGAGATGAGCGGCAAGGCGCAATGAGAGAGAGCTGGGAG 2304
 Qy 720 GluaIagIuAlaIeuhIaIaSerGlyAsnGluIuIaIaProarg---GlyGluCyserGly 738
 Db 2305 GAGGCAAGAGCTCAGAGGCTCTGGGAATGAGTTCCAGAGTTGGGAGTGCAGTGC 2364
 Qy 739 ProaIaThrValaAsnaSerSerAaPthrgIuSerIleProSerProhIsthrgIuaIa 758
 Db 2365 CCAGCTGCTGTCAACAACAGCTCTGATTAAGAGTGCATCCCGGCTTCCAGAGCC 2424
 Qy 759 AlaIyAspThrgIyGlnaengIyProIyPProProaIaThr-----IeuGlyIaAsp 776
 Db 2425 ACGAAGAGACT-----GGGCTTAAACCACTGGCAGTGAACGATTGCCGCTGCC 2475
 Qy 777 GlyProProGluIyProProThrProProArgThrSerArgThraProlIegIuPro 796
 Db 2476 ACCGACCGACCTGTTCT-----CTCCAGAAAGAACCGGACAGCCCTGCTGAGCCC 2529
 Qy 797 ThrProIaSerGluIaIaThrgIyAlaProThrProProProIaIaProProSerProser 816
 Db 2530 TCCCGAGTCCGTGATCCAGTGGCCCAACCATCCCGAGAGCTTCC---CCATCAGCTGGC 2586
 Qy 817 AlaProProProValaIaProIySerGluIuIySerGluIuThraIaIaIaIaProPro 836
 Db 2587 GAACCCCGGCTACTGTGACAAAGATGAACAAGAACCCCGGCTCTCCAGCTCCCGCAG 2646
 Qy 837 ValIgluIyGluIuIySerProProIaIaIaIaGluIuIaIaIaIaIaIaIaIaIaIa 856
 Db 2647 ACAGAAAGATGCCAAGAGAGAGAGTGTGAGGCCAGAGAG-----ATCATGTGGGA 2697
 Qy 857 LysIaIaGluIuPro-----ValIySerGluCythr 867
 Db 2698 AAGCCAGAGAGCCGAGGCTCTGAGAGACCCCGAGAGGTAAAGAGTGACACCAAG 2757
 Qy 868 GluIuIaIaGluIuIyProIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 886
 Db 2758 GAGGAGACCGAGAGAGAGCTGAAAGCAAGCCAGGCGCAGAGGCGCATTTGAACTGTG 2817
 Qy 887 AlaIaGluIyAlaIeuhIyIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 905
 Db 2818 TCTGAGGACCACTTAAAGTGGAG-----GAGGCTGTAGCAAGGACGCTGTGACCAAG 2871
 Qy 906 AlaIySerSerGlyAlaProGlnaSerAaPthrgIuSerIleProSerProhIsthrgIuaIa 925
 Db 2872 GGTTCCAGCTCAGAGTCCAGCCAGAGACAGTGAAGTCCAGTCCAGCTGCAAGTGCAGTAG 2931
 Qy 926 ValaIaSerGluIaIaGluIyGlyAspIyAsnaIyIeuIeuIeuIeuIeuIeuIeuIeu 945
 Db 2932 GTGGAGAAACCCAGAGAGAGTGAAGAGGCGAGCTGTGTACCAAGGCCACCTCTCTC 2991
 Qy 946 ThrProThrgIyAspProaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 965
 Db 2992 ACCCGGCTGGAGATCCCGGCGCAGTACTCCGCCAGAGGCCGCTGAGACTGAAAGCAG 3051
 Qy 966 IeuIyGlnaIyAlaIa 985
 Db 3052 CTGAAGCAGAGAGAGAGCCGCGCATCCCGCATC---GTACCAAGGTCCATGAGCCCGCC 3108
 Qy 986 ATGGIuaAspAlaIa 1005

Db 3109 CCGAGAGACAGTACCCCAAGCCAGTTCCTGTCCTTCCACCCAGAGACCTTA 3168
 Qy 1006 GlnProIySerAaPthrgIuIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 1025
 Db 3169 CAGCCAGAGAGGTGATGTCAGAGTCCGAGAGAAATGCCAGTGGCAAGTCCCGCAGC 3228
 Qy 1026 ProaIaProProIaIaAspIyGlu-----AlaPheIaIaIa 1037
 Db 3229 CCAGTCCCTCTGCGGAGAAAGAGAGAGAAACCCGATTTCTTCCGCTTCCCACT 3288
 Qy 1038 GluaIagIuIyIeuProIyIaPProProCystrPthSerGlyIeuProPheProIa 1057
 Db 3289 GAGGCGCA-AACTACCAAGTGAAGCCCGACCTGTATTCGGCT---GCTTCCAT--- 3342
 Qy 1058 ProProArgIuValIleIySaIaSerProhIaIaIaIaIaIaIaIaIaIaIaIaIaIa 1077
 Db 3343 CTTCCACGGAGAGTATCAAGACTTCAACACCGCT---GACCT---CTTCTTCTTAC 3396
 Qy 1078 AlaProProGluIyIa 1097
 Db 3397 ACACCCCGGATCACCCGCTGCTCTGGGCTTCCAGATAGTGGCCGCGCTCTGCGCA 3456
 Qy 1098 ArgProProThrIleSerAaPthrgIuIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 1117
 Db 3457 GTCCTCCCT---ATCTTAACCCCGACCCCTCATCTCTTCCAGAGATCCCGGCGTA 3513
 Qy 1118 IeuGluArgIuIleGlyAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 1137
 Db 3514 CTGAGAGAGCAGCTGGTGCATCTCCAGGGAGATGTACAGTTCAGTTCGTGTGCTTCA 3573
 Qy 1138 SerGluIa 1157
 Db 3574 TCAGAGCAGCCAG---CCCATGGGCTCTTCCAGATGAGCTGCCCTTCCGCTGAGC 3630
 Qy 1158 ProIyIyIeuAlaIa 1177
 Db 3631 CTTAAGAGCTG----- 3642
 Qy 1178 AlaIyProProGluIuIyIa 1197
 Db 3643 -----GGG 3645
 Qy 1198 ThrAlaIeuGlyIa 1217
 Db 3646 ACAGCACT-GGCTCCGCCACAGTGAAGATCACCAAGGCTT-CCAGTACCCGCGCT 3703
 Qy 1218 ProSerAaPthrgIuIa 1237
 Db 3704 GCAAGAGGCTCC-----AGCTACAGAGCTTATTCACCAAGC-ACGCCCGCAGAGCTC 3756
 Qy 1238 IeuIyIyGlyIa 1257
 Db 3757 CTCTACAGAGGTATACATCAAGAGATCGCGGTGAGAGACAGCCCAAGTGGCTTACCGG 3816
 Qy 1258 GlyArgIuAspSerIeuProIySerIyIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 1277
 Db 3817 GCACAGAGAGACCTGTCACAAAGGCGCATGTATTAAGAGGCAAGAAAGCCACGCTC 3876
 Qy 1278 IeuSerIyGluIyGlyIuIa 1297
 Db 3877 CTATCTATGAAGGTGATGTCTGCTGTACAGTGTCTTAAGAGATGAAGAGAGAGCAGC 3936
 Qy 1298 SerGlyProProhIa 1317
 Db 3937 TCGGGCCACCCCATGAGCTGCGCCCTTAACGCACTTAAGACATGATGAGGCGCGCT 3996
 Qy 1318 ValGlyArgAlaIa 1337
 Db 3997 GTAGGAGAGACTGTACCTCAGCCAGCATGAGAGAGCTATGAGCGCGCCAGC---CCT 4053
 Qy 1338 GluArgIa 1357
 Db 4054 GAGCAGCAGAGCCCT---CACTCAAGAGAGAGATCAATCGAGGCTTCATCAGCGAA 4110

QY 1358 G1Y1LeProArSerTYrValGluAlaGlnGluAspTYrLeuAArgGluAlaLeu 1377
DB 4111 GGATCCCGAGGCTCTATGTGGAGCGAGAGACTTACCGCGGAGCCAGATC 4170
QY 1378 LeuLYaArgGluGlyThrProProProProProSerAArgLeuThGluAlaTyr 1397
DB 4171 TTGAAGGAGAAAGGAGACCACTCCCAACACCACTCGGAGACTGAGAGACTTAC 4230
QY 1398 LyeThrGln-----AlaLeuGlyProleuYsleuYsProAlaHisGluGlyLeu 1414
DB 4231 AAGCCCCGAGCCCTGAGACCTCTGAGTCCCTGAGAGCGAGCTCAGAGAGTGTG 4290
QY 1415 ValAlaThrValIyGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArg 1434
DB 4291 GTAGCACTGTGAAGAGAGCGCGCGCTTATCCATGAGATCCGAGAGAGACTCGC 4350
QY 1435 HisThrProGluLeuProLeuAlaProArgProleuYsGluGlySerIleThrGlnGly 1454
DB 4351 CGCACACTGAGCTACCCCTGGACCAAGGCTCTGAAGAGGTTCCATCACCAGGCG 4410
QY 1455 ThrProleuYsTYrAaPThrGlyAlaSerThrThrGlySerIyGlyHisAaPValArg 1474
DB 4411 ACCCCACTCAAGTACGACTGGGGGACCCCTCACTGGCACCAAGAAACAGAGTGGC 4470
QY 1475 SerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAaPValMetAla 1494
DB 4471 TCCATCTCCGGAGCCCGCGCGCGCTTCCCTGCTCCCTGACCGCGCTGACATATAGCT 4530
QY 1495 AspAlaArgAlaLeuGluAlaGlyCysTYrGluGluSerLeuYsSerAArgProGlyThr 1514
DB 4531 GAGCGCGGGGACACTGAGCGGTGCTGCTATGAAGAGTCTGAAGACCGGTGAGGACC 4590
QY 1515 AlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGly 1534
DB 4591 AGCAGTGTGTCAAGGGGCTCATCACAGTGGGGCTCAGTCGTCGCTGAACTGGGC 4650
QY 1535 LysProArgGlnSerProLeuThrTYrGluAaPHisGlyAlaProPheAlaGlyHisLeu 1554
DB 4651 AAGCCAGCGAAAGCCCACTGACTTACGAAGACCAAGGGGACCCCTTACACAGTCACTG 4710
QY 1555 ProArgGlySerProValIleMetArgGluProThrProArgLeuGlnGlySerLeu 1574
DB 4711 CCAGTGTGCTCCCTGTGACCAAGAGGAGCCACGCGCTTCAAGAGGAGGCTC 4770
QY 1575 SerSerSerIyAlaSerGlnAaPArgIyLeuThSerThrProArgGluIleAlaLys 1594
DB 4771 CTATCCAGAAAGGCTCCAGAACCGGAAGCTGACATCTACACCCCGAGATCGCCAG 4830
QY 1595 SerProHisSerThrValProGluHisHisProHisProIleSerProTYrGluHisLeu 1614
DB 4831 TCCCCACACAGACTGTGCCGAGACACACCTTACCCCATCTCCCTATGACACTTG 4890
QY 1615 LeuArgGlyValSerGlyValAaPLeuTYrArgSerHisIleProLeuAlaPheAaPPro 1634
DB 4891 CTCGGGGGCTGACTGTGTGACCTGTACCGTGTACATCCATTGGGCTTTGACCCC 4950
QY 1635 ThrSerIleProArgGlyIleProLeuAaP---AlaAlaAlaAlaTYrTYrLeuProArg 1653
DB 4951 ACCTTCATACCCGAGGAGTCCCTCTGAAAGACGACCCGAGCTTACTCTGCCCGG 5010
QY 1654 HisLeuAlaProAaPProThTYrProHisLeuTYrProProTYrLeuIleArgGlyTYr 1673
DB 5011 CACTGGCCCCCGACCCCACTTACCACTGTACCTTACCTTACCTTACCTTACCTTAC 5070
QY 1674 ProAaPThrAlaAlaLeuGluAaPArgIleThIleAaPArgTYrIleThSerGln 1693
DB 5071 CTGACACGGGCGGCTCGAGAACCGGCAACCATCATATGATCATCATCTGAC 5130
QY 1694 GluMetHisHisAaPThrAlaThrAlaMetAlaGlnArgAlaAaPMetLeuArgGlyLeu 1713
DB 5131 CAGATGACACAAAGCTGCTCGCATGAGCCCAAGGTGCTGACATGTGAGGGGTCTG 5190

QY 1714 SerProArgGlnSerSerLeuAlaLeuAaPThrAlaAlaProArgGlyIleIleAaP 1733
DB 5191 TCACCGAGAGATCTCTGCTGGCCCTCAATTATTCGCTGGCCCAAGAGCATTATGAC 5250
QY 1734 LeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAlaThr 1753
DB 5251 CTGTCCCAAGTCCACACCTGCGCTGTGTGACCAACGCGAGGACCCCTGCCAC 5310
QY 1754 AlaMetAaPArgLeuAlaTYrLeuProThrAlaProGlnProPheSerSerArgHisSer 1773
DB 5311 GCATGACAGGCTTGTGCTTACCTTCCACTGCGGCCCACTTTCAGAGCGCCACAGT 5370
QY 1774 SerSerProLeuSerProGlyGlyProThrHisLeuThTYrProThrThThSerSer 1793
DB 5371 AGCTCACGGCTGTCCAGAGAGGCCCACTTCACTTAAACCACTGACCATCTTCA 5430
QY 1794 SerGluArgGluArgAaPArgAaPArgGluArgAaPArgAaPArgGluArgGluLysSer 1813
DB 5431 TCGGAGCGGGAACGGAACTGAGCGGGAACGAGAC-----AAGTCC 5472
QY 1814 IleLeuThrSerThrThrThrThrValGluHisAlaProIleTPraArgProGlyThrGln 1833
DB 5473 ATCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 5532
QY 1834 SerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerSerAArgProAlaSer 1853
DB 5533 AGCAGCGGGCT-----GGGGGACGACCCCGCGCTCC 5568
QY 1854 HisSerHisAlaHisGlnHisSerProIleSerProArgThrGluAaPAlaLeuGln 1873
DB 5569 CACAC-----CACGACACTGCGCCATCTCCCGCGACCGACGCGCTTGCACAG 5622
QY 1874 ArgProSerValLeuHisAaPThrGlyMetLysGlyIleIleThrAlaValGluProSer 1893
DB 5623 AGCGCCAGTGTGTGCAACACGAGCATGAAGGCTGTGTCACTCGTGAACCCGCG 5662
QY 1894 LysProThrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaThrPhe 1913
DB 5683 AGCGCCAGGCTCTGAGTGTCACTTCCCTTCCGCTTCCGCGCAGTGCACATTC 5742
QY 1914 ProProAlaThrHisCysProLeuGlyGlyThrLeuAaPArgIyValTYrProThrLeuMet 1933
DB 5743 CCACCTGCCACCACTGCGCCACTTGTGTGACACCTTGAAGGGGTCTACCTACCTCATG 5802
QY 1934 GluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProAla 1953
DB 5803 GAGCCGCTCTGTACCAAGAGACCTCTGAGTGTGCGCGCGCGAGCGCGCGGTG 5862
QY 1954 AspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSer 1973
DB 5863 GACGATGCGATGCTTCTTCTCACCAACCCCGCGCG-----GAGCCGCTCC 5913
QY 1974 SerProSerIyGlySerGluProArgProLeuValProProValSerGlyHisAlaThr 1993
DB 5914 TCACCCAGCAAGAGCTCCGAGCCCGATCCCTAGACCCCCAGAGTCCACACACAGCC 5973
QY 1994 IleAlaArgThrProAlaLysAaPLeuAlaProHisHisAlaSerProAaPProProAla 2013
DB 5974 ATGCGCGGACCCCAAGAGCTTGTGACCCCAACATCCAGTCCGAGCCGCGCGGG 6033
QY 2014 ProProAlaSerAlaSerAaPProHisArgGluLeuThGlnSerIySProPheSerIle 2033
DB 6034 ---CCACCTCGGCTCAATCTGCACCGAAGAAAGCTTAAGTAACCTTTTCATC 6090
QY 2034 GlnGluLeuGluLeuArgSerLeuGlyTYrHis---GlySerSerTYrSerProGluGly 2052
DB 6091 CAGGAATTGAACTCGTCTCTGAGTACCAAGAGAGCTGCTTACACCCCGATGGG 6150
QY 2053 ValGluProValSerProValSerSerProSerLeuThHisAaPArgIyLeuProLys 2072
DB 6151 GTGAGCGCATGAGCGCGGTGAGCTCCCGAGCTTACCAACCAAGAGGGGCTCTCCAA 6210
QY 2073 HisLeuGluGluLeuAaPArgSerHisLeuGluGlyGluLeuArgProLysGlnProGly 2092

QY 101 GluPhe11eGluSer1ySarProArgLeuGluLeuProAspProLeuLeuArgPro 120
Db 460 GAATTACCGAGGAGCAAGCGCCCGCTGGAGCTACTACCGATACCCCTGCGGCCA 519
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLyAspArgSer 140
Db 520 TCACCCCTGCTGGCCACTGGGACCGAGGTGCTGMAACCTTACCAAGGACCGTAGC 579
QY 141 LeuThG1yLyLeuGluProValSerProProSerProProh1SerhAspProGluLeu 160
Db 580 CTGGGAGGAGAGGAGCTGTGTCTACCTCCAGTCCCGACCGCTGACCTTGAAGCTA 639
QY 161 GluLeuVal1ProProArgLeuSerLyGluGluLeu11eGluAsn---MetAspArgVal 179
Db 640 GAGCTGGCCCATCTGCAGCTGTCCAAAGAGAGACTGATCCAGAACAGATTGACCCGCTG 699
QY 180 AspArgGlu11eThrMetValGluGlnGln11eSerLyLeuLybLybGlnGlnGln 199
Db 700 GACCGTGAGATCACCACTGGTAGAGCAGCATCTCCAGCTGMAAGAAAGACGAAACAG 759
QY 200 LeuGluGluGluAla1Ala1ySarProProGluProGluLybProVal1SerProPro11e 219
Db 760 TTGGAGAGAGAGGCGCCCAAGCGCCGAAACCGAGAGAGCTGTGTGCGCACCAACATA 819
QY 220 GluSerLyb1ySarSerLeuValGln11e1e1yTAspGluAsnArgLybLybAlaGlu 239
Db 820 GAATCAAAAGCACCGAAGCCTGTGCTCAGATCATCTTACATGAGAACCGGAAGAAAGCGAA 879
QY 240 AlaAla1ySarG11eLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyTAsnGln 259
Db 880 GCCGACACCGGATCTTAAAGGCTGGGGCCCAAGTGGAGCTGCTGTCAACACAG 939
QY 260 ProSerAspThrArgGlnTyTh1yGluAsn11eLyb11eAsnGln11eMetArgLybLyb 279
Db 940 CCGTCTGACACACGCAAGTACCATGAAACATCAATAATAAACAGGCGAAGCGGAAAG 999
QY 280 Leu11eLeuTyTh1yThLybSarArgAsn11eAlaArgLybGlnTyThLybGlnTybPheCys 299
Db 1000 CTGATCTTCTTACTTTTAAAGGAGGAACCAACGCGGCAAGCATGAGGAACAGCCTTCTGC 1059
QY 300 GlnArgTyTAspGlnLeuMetGluAlaLeuGluLybLybValGluArg11eGluAsnAsn 319
Db 1060 CAGCGCTATGACACGCTCATGTGAGCGTGGAGAAAGATGAGAGCGCATAGAGAAACAAT 1119
QY 320 ProArgArgArgAla1ySerGluSerLybValArgGluTyThTyThGluLybGlnPheProGlu 339
Db 1120 CCGGAAAGAGGAGGCAAGAGAGCAAGTGAAGGAGTACTACAGAAACAGTTCCCGAG 1179
QY 340 11eArgLybGlnArgGluLeuGlnGln11eArgMetGlnSerArgValGlyGlnArgLybSer 359
Db 1180 ATCCGCAAGCACGCGGAGCTGACAGAGCCCATGCAAGACAGGCTGGCCACGCTGGCACT 1239
QY 360 GlyLeuSerMetSerAla1AlaArgSerGlu11eGluValSerGlu11e11eAspGlyLeu 379
Db 1240 GGGCTCTTCATGTCGGCTCCCGCAGTGAAGCATGAGGTTTCTGAGATCATTTGAGCTTG 1299
QY 380 SerGluGlnGluAsnLeuGluLybGlnMetArgGlnLeuAlaVal11eProMetLeu 399
Db 1300 TCTGAGCAGAGAACTTGAAGAGAGATGCGCAGCTGCGCTGATCC--GCCATGTG 1357
QY 400 TyTAspAlaAspGlnGlnArg11eLybPhe11eAsnMetAsnGly1yLeuMetAlaAspPro 419
Db 1358 TACGAGCGC--GACCAAGAGAGATCAAGTTCAATCAATGATGAGACTCATGTGATGACCC 1416
QY 420 MetLybValTyThLybAspArgGlnValMetAsnMetTTPSerGluGlnGluLybGluThr 439
Db 1417 ATGAAGGCTTACAAAGACCGTCAAGTTTACCAATCTGAGAGAGAGAGAGGAGGAGCAC 1476
QY 440 PheArgGluLybPheMetGln11eAspProLybAsnPheGlyLeu11e11eAspPheLeuGlu 459
Db 1477 TTCCTGTAGAAAGTTTATGACAGCACTTAAAGAACTTTGGCCTGATTCCTTCTTGAG 1536
QY 460 ArgLybThValAlaGluCysValLeuTyThTyThLeuThLybLybAsnGluAsnTy 479

Db 1537 AGAAAGACGTCGCTGAGTGTGTCTCTATTAATTAATCACTGACCAAGAAATGAATGAATAC 1596
QY 480 LybSerLeuValArgArgSerTyThArgArgGlyLybSerGlnGlnGlnGlnGln 499
Db 1597 AAGAGCTTGATGAGGAGAGCTTATCGGCCCGGTGAAGAGCCAGAGAGAGAGAGAGAG 1656
QY 500 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 519
Db 1657 CAACAAACAGCAG 1704
QY 520 LybAspGluLybGluLybGluLybGluAlaGluLybGluGluLybProGluValGlu 539
Db 1705 AAG 1764
QY 540 AsnAspLybGluAspLeuLeuLybGlybTyThArgAspThSerGlyGluAspAsnAsp 559
Db 1765 AACGAG 1824
QY 560 GluLybGluAlaValAlaSerLybGlyArgLybThAlaAsnSerGlnGlyArgArgLyb 579
Db 1825 GAGAAAG 1884
QY 580 GlyArg11eThrArgSerMetAlaAsnGlnAlaAsnSerGluGluAla11eThrProGln 599
Db 1885 GGCCTATACCGGCTCCATGGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1944
QY 600 GlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTyThArgGluGlu 619
Db 1945 CAAGTTTCAAGCTGTGCTTCAATGAGATGAACAGAGATTCCTGCTGAGCTGAGAGAG 2004
QY 620 MetGluThAlaLybLybGlyLeuLeuGln11eLybLybMetGluLybGluArgGln 639
Db 2005 ATGAGACACGAAAG 2064
QY 640 MetValGlySerLybThValSerGlnCysLybAsnPheTyThPheAsnTyThLybLybArg 659
Db 2065 ATGTGTGGCTCCAAAGCGGTGTCCAGGTGAAGAACTTCACTTCACTTCACTTCACTTCA 2124
QY 660 GlnAsnLeuAspGlu11eLeuGlnGln11eLybLybLeuLybMetGluLybGluArgGln 679
Db 2125 CAGAACTTGAAGAAATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2184
QY 680 ArgArgLybLybLybLybLybAlaProAlaAla1ySarGluGluAla1ySarProVal 699
Db 2185 CGAG 2244
QY 700 ValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGlu 719
Db 2245 GCTGAG 2304
QY 720 GluAlaGluAlaLeu11eAlaSerGlybAsnGluValProArg---GlyGluCysSerGly 738
Db 2305 GAGGAG 2364
QY 739 ProAlaThrValAsnAsnSerSerAspThArgSerLybProSerProh1SerhGluAla 758
Db 2365 CAGGCTGCTTAAACAAGCTCTGATCTGAGAGTGTCCATCCCGGCTTCAAGAGCC 2424
QY 759 AlaLybAspThGlyGlnAsnGlyProLybProProAlaThr-----LeuGlyAlaAsp 776
Db 2425 ACGAAG 2475
QY 777 GlyProProProProGlyProProProProProProProProProProProProProPro 2529
Db 2476 ACCAG 2586
QY 797 ThrProAlaSerGluAlaThrGlyAlaProThProProProProAlaProProSerProSer 816
Db 2530 TCCCAAGTCCCGAAG 2586
QY 817 AlaProProProVal1yProLybGluGluLybGluGluGluThThAla1Ala1yProPro 836

Db	2567	GCACCCCGGCTACTCTGTGACAGATGACAGAAAGCCCGGCTCTCCAGCTCCCAAG	2646
OY	837	ValGluGluGluGluGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGly	856
Db	2647	ACAGAAATGTCACCAAGAGACAGAAAGTCTGAGGCCGAGAG-----ATGATGTGGGA	2697
OY	857	LysAlaGluGluPro-----ValLysSerGluCysThr	867
Db	2698	AAGCCAGGAGAGCCCGAGAGCCTCTGAGAGCCCGGAGAGTGTAAAGTACCAAG	2757
OY	868	GluGluAlaGluGluGluGlyProAla---LysGlyLysAspAlaGluAlaAlaGluAlaThr	886
Db	2758	GAGAGAGACCGAGGAAGACCTGTAAGACAAAGCAAGGGCACAGAGCCATTGAACCTGTG	2817
OY	887	AlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyAlaGlyAlaThrThr---	905
Db	2818	TCTGAGGACACCTTAAGTGGAG-----GAGGCTGTATCAAGAGCGCTGTGACCAAG	2871
OY	906	AlaLysSerSerSerGlyAlaProGlnAspSerLysSerSerAlaThrCysSerAlaAspGlu	925
Db	2872	GGTTCACACTCAGGCGCACCCAGGACAGTACTCAGCTGCACCTCGAGTGCAGTAG	2931
OY	926	ValAspGluAlaGluGlyGlyLysAspLysAsnArgLeuLysSerProArgProSerLeuLeu	945
Db	2932	GTGACAGAACCCGAGAGAGTGCACAGGGCAGGCTGCTGTACCAAGGCCAGACTCTTC	2991
OY	946	ThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGln	965
Db	2992	ACCCGGGCTGGAGATCCCGGGCCAGTACCTCGCCCAAGAGCGCTGACTGAAGAG	3051
OY	966	LeuLysGlnAlaGluAlaAlaAlaLeuProProIleGlnValThrLysValHisGluProPro	985
Db	3052	CTGAAGCAGCGAGCGCGCCATCCCCCTATC---GTCCACAGGTCCATGAGGCCCC	3108
OY	986	ArgGluAspAlaAlaProThrLysProAlaProProAlaProProProGlnAlaLeu	1005
Db	3109	CGGGGAGACAGTACCCCCANAGCCAGTTCCTCTGCTCCACCCAGCAGACCTTA	3168
OY	1006	GlnProGlnSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSer	1025
Db	3169	CAGCCAGAGGGGTGAGTGTCTCAGCAGTCCGAGAGAAATCCACGTGGGACGTCCCGCAGC	3228
OY	1026	ProAlaProProAlaAspLysGlu-----AlaPheAlaAla	1037
Db	3229	CCAGTGCCTCTCCGCGAAGAGGACAGAAACCCGATTCCTTCCGGCTTCCCACT	3288
OY	1038	GluAlaGlnLysLeuProGlyLysProProCysTrpThrSerGlyLeuProPheProAla	1057
Db	3289	GAGGGCCA-AGCTAACCGACTAGGCCCCACCGCTGTATCATGGCT--GCCCTCCAT--	3342
OY	1058	ProProArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyr	1077
Db	3343	CCTCCACGGGAGGTATCAAGACTTCCACAGCGCT---GACCT--CTCTTCTCTAC	3396
OY	1078	AlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuPro	1097
Db	3397	ACACCCCGGTCACCCGCTGCTCTGGGCTCCACAGTAAAGCGCGGCCGCTCTGGCA	3456
OY	1098	ArgProProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerVal	1117
Db	3457	CGTCCCCC---ATCTTCAACCCCAACCCCTCATCTCTGCGCAACATCCCGCGCTA	3513
OY	1118	LeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValaProTyr	1137
Db	3514	CTTGAGAGCGACTGTGGTCCATCTCCAGGGGATGTACATCCACCTTCGTGTGCTAC	3573
OY	1138	SerGlnHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAsp	1157
Db	3574	TCAAGAGCATGCCAAG---CCATAGGCCCTCTCACATGAGAGTCCCTTCGTCGGTGAG	3633
OY	1158	ProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGln	1177
Db	3631	CCTAAGAAGCTG-----	3642

QY	1178	AlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGly	1197
Db	3643	-----Ggg	3645
QY	1198	ThrAlaLeuGlySerValProGlyGlySerIleThrIlyGlyIleProSerThrArgVal	1217
Db	3646	ACAGCACT-GGCTCGGCACACAGTGGAAAGCATCAACAAGGGCTT-CCAGTACCCGGGCT	3703
QY	1218	ProSerApeSerAlaIleThrTyArgGlySerIleThrIleGlyThrProAlaAspVal	1237
Db	3704	GCAGACGGCCCC-----AGCTACAGAGGCTCTTACACCAACGC-ACGCCCGAGAGCTC	3756
QY	1238	LeuTyIlyrsGlyThrIleThrArgIleIleGlyIleuAspSerProSerArgLeuAspArg	1257
Db	3757	CTTACAAAGGATCAATCAACAGAGATGCTCGTATGAGACACCCCAAGTCCGCTTGACCG	3816
QY	1258	GlyArgGluApeSerLeuProGlyIleValIleTyrgGlyIlyIlyIlyVal	1277
Db	3817	GCACGAGAGACACCCCTGCCCCAAGGCCATGCTATATGAGGCGCAAGAAAGCCACGTC	3876
QY	1278	LeuSerTyrgIlyGlyIlyMetSerValThrGlnGlySerIlyGlyIlyuAspArgIlySerSer	1297
Db	3877	CTATCTATGAAAGGATGATGCTCGGTGCACAGTCTCTAAGGAGATGGAAAGACAC	3936
QY	1298	SerGlyProProHISgluThrAlaAlaProIlyAspThrTyraPwMetGlyIlyArg	1317
Db	3937	TGGGGCCACCCCATGAGACTGCCGCCCTTAAACGACCTATGATCATGTGAAGGCGGT	3996
QY	1318	ValGlyArgAlaIleSerSerAlaSerIleGlyIlyLeuMetGlyIlyArgAlaIleProPro	1337
Db	3997	GTAGCGAGAGTATGCACCTCGACAGATAGAGAGACTCATAGGGCCGCGCATC--CCT	4053
QY	1338	GluArgHISerProHISleuIlyGlyGlnIleHISleArgGlySerIleThrGln	1357
Db	4054	GAGCAGCACAGCCCC--CACTTAAGAGAGAGATCATATCCAGGCTCATACCGAA	4110
QY	1358	GlyIleProArgSerTyraValGluAlaGlnIleuAspTyreIleuArgArgGluAlaIlyLeu	1377
Db	4111	GGCATCCCGAGTCTCTATGTGAGGCGCAGAGAGACTTATCCGCGGAGGCCAAGCTC	4170
QY	1378	LeuIlyAspArgIlyGlyThrProProProProProSerArgAspLeuThrGluAlaTyr	1397
Db	4171	TTGAAGCCAGAAAGGACACACCTCCCCACACACACTCGGACCTGACTGAACTATAC	4230
QY	1398	LysThrGln-----AlaLeuGlyProLeuIlySerProAlaHISgluGlyLeu	1414
Db	4231	AAGCCCGGCCCCCGAGACCTCTGGGTCCCTGAAGCTGAAGCGGACTCAGAGGGGTGTG	4290
QY	1415	ValAlaThrValIlyGlyAlaGlyIlyArgSerIleHISgluIleProArgGluIleuArg	1434
Db	4291	GTAGCAACTGGAAGAGAGCGCGCGCTCTATCATGAGATCCGGAAGAGAGACTCGC	4350
QY	1435	HISThrProGluIleuProLeuAlaProArgProLeuIlyGlySerIleThrGlnGly	1454
Db	4351	CGACACACTGAGCTACCCCTCGACACACAGGCTCTGAAGAGAGGTTTCATCATCCAGGCG	4410
QY	1455	ThrProLeuIlyTyraPbThrGlyAlaSerThrThrGlySerIlySHISAspValArg	1474
Db	4411	ACCCCACTCAAGTACGACTCTGGGGCACCCCTCACTGGCACCCGCTGGAATATAGCTCGC	4470
QY	1475	SerLeuIleGlySerProGlyIlyArgThrPheProProValHISProLeuAspValMetAla	1494
Db	4471	TTCATCATCGGCAGCCCGCGCGCTTTCCTCCCTCGACCCGCTGGAATATAGCTCGCT	4530
QY	1495	AspAlaArgAlaLeuGluArgAlaCysTyrgIlyGluIleSerLeuIlySerArgProGlyThr	1514
Db	4531	GACGCCCGGCACTGAGCGTGCCTCGTATGAAGAGTCTGAAGAGCCGCTCAGGAC	4590
QY	1515	AlaSerSerSerGlyIlySerIleAlaArgGlyAlaProValIleValProGluIleGly	1534
Db	4591	AGCATGTGTGAGGGGCTTCATACACGTGGGGCTCCAGTGTGTGTGCTCAACTGGGC	4650

1535 LysProAlaGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisGlyLeu 1554
1555 ProAlaGlnSerProAlaThrMetArgGluProThrProArgLeuGlnGlySerLeu 1574
1575 SerSerSerIleAspArgIleAspArgIleAspArgIleAspArgIleAspArgIleAspArg 1594
1595 SerProHisSerThrValProGluHisIleProHisProIleSerProTyrGluHisGlyLeu 1614
1615 LeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspPro 1634
1635 ThrSerIleProArgGlyIleProLeuAsp---AlaAlaAlaIleTyrTyrLeuProArg 1653
1654 HisLeuAlaProAsnProThrTyrProHisGlyLeuTyrProProTyrLeuIleArgGlyTyr 1673
1674 ProAspThrAlaAlaLeuGluLeuAsnArgIleThrIleIleAsnAspTyrIleThrSerGln 1693
1694 GlnMetHisIleAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeu 1713
1714 SerProArgGluSerSerLeuAlaLeuAsnTyrAlaIleGlyProArgGlyIleIleAsp 1733
1734 LeuSerGlnValProHisLeuProValLeuValProProThrProGlyLysProAlaThr 1753
1754 AlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSer 1773
1774 SerSerProLeuSerProGlyGlyProThrHisGlyLeuThrLysAspArgGluArgGlyLysSer 1793
1794 SerGluArgGluArgAspArgAspArgGlyLysArgAspArgAspArgGluArgGlyLysSer 1813
1814 IleLeuThrSerThrThrThrValGluHisAlaProIleThrAspProGlyLysGln 1833
1834 ATCTCAAGCTTCAACATCACTGAGACATGCAACCATCTGAGACCTGAGACGAG 5532
1834 SerSerGlySerSerGlySerSerGlyGlyGlySerSerSerArgProAlaSer 1853
1854 HisSerHisIleAlaHisGlnHisSerProIleSerProArgTyrGlnAspAlaLeuGlnGln 1873
1874 ArgProSerValLeuHisIleAsnThrGlyMetLysGlyIleIleThrAlaValGluProSer 1893
1894 LysProThrValLeuArgSerThrSerThrSerSerProValArgProAlaIleThrPhe 1913

5683 ACCGCCAGGCTCTGAGGTCACCTCACCTCTCCCTGTCCGCCAGCTGCACATTC 5742
1914 ProProAlaThrHisIleCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMet 1933
5743 CACCTGACCAACCACTGCGCCACTTGGGACCTTGAAGGGGTCTACCTTACCTCATG 5802
1934 GluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProArgAla 1953
5803 GAGCCGCTCTGTATCCAAAGAGACCTCTCGGGTGGCCGGCCGAGCGGCCCGGTG 5862
1954 AspThrGlyHisIleAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSer 1973
5863 GACGATGGCCATGCTCTTCTCCACCAAAACCCCGCGCGG-----GAGCCGCGCTCC 5913
1974 SerProSerGlySerGluProArgProLeuValProProValSerGlyHisIleAlaThr 1993
5914 TCACCCAGAGAGAGCTCCAGACCCGATCCCTAGACACCCCGCTTCAGCCACAGCC 5973
1994 IleAlaArgThrProAlaLysAsnLeuAlaProHisIleAlaSerProAspProProAla 2013
5974 ATGCGCCGACCCCGCAAGAGACCTTGCACCCCAACATGCCAGTCCGACCGCGGG 6033
2014 ProProAlaSerAlaSerAspProHisIleArgGlyLysThrGlnSerLysAspPheSerIle 2033
6034 ---CCCACTCGGCTCAGATCTGACACGAGAAAGACTCAAAAGTAACCTTTCCATC 6090
2034 GlnGluLeuGluLeuArgSerLeuGlyTyrHis---GlySerSerTyrSerProGluGly 2052
6091 CAGGAATTGAATCTCGTCTCTGGGTTCACAGCTGAGCTGGCTTACAGCCCGGATGG 6150
2053 ValGluProValSerProValSerProSerProSerLeuThrHisAspLysGlyLeuProLys 2072
6151 GTGAGGCCATGACCGCGGTGAGCTCCCAAGCTTACCCACAGCAAGGGGCTCTCCAA 6210
2073 HisLeuGluGluLeuAspLysSerHisGlyLeuGluGlyLysLeuArgProLysGlnProGly 2092
6211 CCTCTGAAGAGACTGAGAAAGAGCACTTGAAGGGAGACTCGGACCAAGACCGAGGC 6270
2093 ProValLysLeuGlyGlyGluAlaHisIleAspProHisGlyLeuArgProLeuProGlySer 2112
6271 CCCATGAGCTAGCGCGGAGCTGCGCATCTCCACATCTCGGCACTGCGCGGAGG 6330
2113 GlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgVal 2132
6331 CAGCCCTATCAAGCCCACTCTCCAGACTGCGCCAGGATCAAGAGTACCAAGAGGTG 6390
2133 ValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisIlePro 2152
6391 GTACACCTGGCTCAGACATCAGCAGAGTCATTAGCAGAGACTACACCGGACCAACCG 6450
2153 GlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyLysAspCysPro 2172
6451 CAGCAGCTTAGAGGCCCTTCCGCTCTCTACTCTCTTCCGAGCCAGCTGCGCT 6510
2173 ValLeuAspLeuAlaArgProProSerAspLeuTyrLeuProProProAspHisGlyAla 2192
6511 GTGCTGATCTTCCGCGCCAGCACTGACCTTACCTCCACCCCGCAACATGAGCAC 6570
2193 ProAlaArgGlySerProHisSerGluGlyLysArgSerProGluProAsnLysThr 2212
6571 CCAGCCCGGAGTCCCGCCACAGTGAAGGGGCAAAAGGTCCCGAGAACCCAGCAAAACA 6630
2213 SerValLeuGlyGlyGlyGluAspArgIleGluProValSerProProGluGlyMetThr 2232
6631 TCGATCTCGGAGAGAGTAGAGATCCATTAGCTGTGTCCCAACAGAGGACATGACT 6690
2233 GluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThr 2252
6691 GAGCCAGAGACATGCTCGGAGGCTGTATCCACATGCTGTATCGAGAGGGGACAGGCG 6750
2253 GluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePhe 2272

```
Db 6751 GAGCCC--AGAGTGGGCTCTAAGTCTCCAGGCAACACAGCCAGCCGCGACCTTCTTC 6807
Oy 2273 SerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnLysLeuLys 2292
Db 6808 AGTAAGCTGACTGAGAGCACTCCGCCATGCTGTAAGTCTGAGAGACAGAGATCAACAAG 6867
Oy 2293 LysLeuAnthrHisAsnArgAsnGluProGluTyrAsnLysSerGlnProGlyThrGlu 2312
Db 6868 AAACTCACACCCACACACCGGAGCAGACAGAAATACATATTATGGCAGCTGGAGCGGAA 6927
Oy 2313 IlePheAnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaVal 2332
Db 6928 ATCTTCACATGCGCCCATCATCTGAGCAGAGCGCTTATGACTGTAGAAGCCAGCGGTG 6987
Oy 2333 GlnGluHisAlaSerThrAsnMetGlyLeuGlnAlaIleLeuArgLysAlaLeuMetGly 2352
Db 6988 CAGAACACGCGCAGACCAACATGGGGCTAGAGGCCATTATTGAAAGCACTCATGGGT 7047
Oy 2353 LysTyrAspGlnTTPGlnGluSerProProLeuSerAlaAsnAlaPheAnProLeuAn 2372
Db 7048 AAATATGATCAGTGGAGAGAGCCCGCGCTCGGCGCAATGCTTTAACCTCTGAAT 7107
Oy 2373 AlaSerAlaSerLeuPro---AlaAlaMetProIleThrAlaAlaAspGlyArgSerAsp 2391
Db 7108 GCCAGCGCCAGCTGCGCCGCTGCTCTATGCCATTAACCACTGTAGCGAGCGAGGTAGC 7167
Oy 2392 HisThrLeuThrSerProGlyGlyGlyLysValLysValSerGlyArgProSerSer 2411
Db 7168 CACGACCTCACCTCGCCAGGTGGAGTGGAGAAAGCCAAAGTCTCTGGCAGACTTAGCAGC 7227
Oy 2412 ArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSer 2431
Db 7228 CGAAAGCCCAAGTCCGCCAGACCAAGCCCTAGCGCTCGGAGACCGACCCCTCTCTCTCC 7287
Oy 2432 SerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgAlaThrGlu 2451
Db 7288 TCAGTACACTCAGAGGGGAGTGCATGCGCGAACAACACTCAACCGGTGTGGAGAG 7347
Oy 2452 AspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeu 2471
Db 7348 GACCGGCCCTCATCTGCGAGGGTCCACGCCATTCCCTCAACCCCTTTGATTATGAGGCTA 7407
Oy 2472 GlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeu 2491
Db 7408 CAGGAGAGGTGATGAGCTCCCGGCCCACTGGCCTTGGCAGGCGAGCGGCCCTCA 7467
Oy 2492 AlaGlyProHisAlaIleThrAspGluProLysProLeuLeuCysSerGlnTyrGlu 2511
Db 7468 GCTGGTCCCAACAGCGCTGGGATGAGAGCCCAAGCCACTGCTGTTCACAGTATGAG 7527
Oy 2512 ThrLeuSerAspSerGlu 2517
Db 7528 ACACCTCTCGGACAGCGAG 7545

RESULT 13
US-10-087-192-650
; Sequence 650, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 650
```

```
; LENGTH: 7534
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-087-192-650

Alignment Scores:
Pred. No.: 0
Score: 9691.50
Percent Similarity: 87.70%
Best Local Similarity: 84.06%
Query Match: 73.34%
DB: 14 Gaps: 28

US-09-522-753-5 (1-2517) x US-10-087-192-650 (1-7534)
Oy 295 LysGlnLysPheCysGlnArgTyrAspGlnLeuMetGlnAlaLeuGluLysValGlu 314
Db 2 GACACGCCCTTCTGCCAGCGCTATGACCACTCATGAGCGGTGGAGAAAGATAGAG 61
Oy 315 ArgIleGluAsnAsnProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGlu 334
Db 62 CGCATAGAGAACAAATCCCGAAGAGGCGCCAAAGAGAGCAAGTGAAGATCACTACAG 121
Oy 335 LysGlnPheProGluIleArgLysGlnArgGluLeuGlnGlnLysArgMetGlnSerArgVal 354
Db 122 AAACAGTTCGCGAGATCCGACAGCAGGAGCTGACGAGCGCATGACAGAGCGGTG 181
Oy 355 GlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGlu 374
Db 182 GGCACGCTGGCAGAGGGGCTCTCCATGTCGCTGCCCGCCAGTGAAGCAAGAGTTTCTGAG 241
Oy 375 IleIleAspGlyLeuSerGlnGlnLysLeuGlnLysGlnMetArgGlnLeuAlaVal 394
Db 242 ATCATTTAGTGGCTGTGTGACAGAGAACTGGAGAAAGATGCGCAGCTGGCGGTG 301
Oy 395 IleProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGly 414
Db 302 ATCCCGCCCATGTTGTATGACGCGGACGACGAGAGATCAAGTTATCAATGAATGA 361
Oy 415 LeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTyrSerGlu 434
Db 362 CTCATGATGATGACCCCATAGAGTCTCAACAGACCTCAGGTTACCAATGTGAGCGAG 421
Oy 435 GlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIle 454
Db 422 CAGAGAGAGACACCTTCGTAAGAACTTATGACGACCCCTAAGAACTTTGGCCTGATT 481
Oy 455 AlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLys 474
Db 482 GCCTCATTTCTGAGAGAGAAAGACGGTGGTGAAGTGTCTCTTATTACTTACCTGACAA 541
Oy 475 LysAsnGluAsnTyrLysSerLysLeuValArgArgSerTyrArgArgArgGlyLysSerGln 494
Db 542 AAGAAATGAATTTACAAAGACTTGGTGAAGCGGAGACTTACCGCGCGGTGGCAAGCGCAG 601
Oy 495 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 514
Db 602 CAGCAGCAGCAGCAGCAACAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 649
Oy 515 SerSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 534
Db 650 AGCAGCCAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 709
Oy 535 LysProGluValAlaAsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSer 554
Db 710 AAGCAGAGATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 769
Oy 555 GlyLysAspAsnAspGluLysGlnAlaValAlaSerLysGlyArgLysThrAlaAsnSer 574
Db 770 GCGCAGAGCAACGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 829
Oy 575 GlnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGlu 594
```


[illegible]

QY	941	ArgProSerLeuLeuThrProThrglyAspProArgLysAlaMetLysProGlnLysPro	960
Db	1937	AGGCCCAAGCTCTTCAACCCGAGCTGGAGATCCCGGGCCAGTACCTGGCCCAAGACCG	1996
QY	961	LeuAhpLeuLysGlnLeuLysGlnArgLysAlaAlaLysProIleGlnValThrLys	980
Db	1997	CTGAGCCTGAAGCAGCTGAAGCAGCGAGCGCGCATCCCTTATC---GTCAACAAAG	2053
QY	981	ValHisGluProProArgGluAspArgLysAlaProThrLysProAlaProProAlaProPro	1000
Db	2054	GTCCATGAGGCCCCCGGAGAGACACGTACCCCAAGGCAATTCCTCCCTGTGGCTTCA	2113
QY	1001	ProProGlnIleuLeuGlnProGlyLysSerAspAlaProGlnGlnProGlySerSerProArg	1020
Db	2114	CCCAAGCAGACACCTACACGCGCAGAGGGTGAAGTGTCTTCAGAGTCCGGAGGAAATCCACGT	2173
QY	1021	GlyLysSerArgSerProAlaProProAlaLysAspLysGlu-----	1033
Db	2174	GGCAAGTCCCGCAGCCAGTGGCTCTCCGCGAAGAGGACAGAAACCCGCAATCTTT	2233
QY	1034	---AlaPheAlaIleGlnValGlnLysLeuProGlyAspProProGlySerThrTrsGly	1052
Db	2234	CCGGCTTTTCCCACTGAGGGCCCAAGACTACCACTGAGCCCCACGCTGTGATATGGGC	2293
QY	1053	LeuProPheProValProProArgGluValIleLysAlaSerProHisAlaProAspPro	1072
Db	2294	CTGGCCTTTCCCAATCCTCCACGGGAGGTATCAAGACTTCCCAACACGCGGTGACCC	2353
QY	1073	SerLysPheSerLysAlaProProGlyHisAspLeuProLeuGlyLeuHisAspThrAla	1092
Db	2354	TCTGCTCTTCTCTACACACCCCGGTCACCCGCTCTCTGGGCTTCCACAGTAACTGGC	2413
QY	1093	ArgProValLeuProArgProProThrLysSerAspProProProLeuLysSerAla	1112
Db	2414	CGGCGCTCTCTGACAGCTCCCC---ATTCTTAACCCCCACCTCTATCTCTGGC	2470
QY	1113	LysHisProSerValLeuGlnArgGlnIleGlyValIleSerGlnGlyMetSerValGln	1132
Db	2471	AAGCATCCCGGCGGACTTGAGAGGAGCGTGGGTGCATCTCCACGGGAGTGTACGTCCAG	2530
QY	1133	LeuHisAlaValProLysSerGlnHisAlaLysAlaProValGlyProAlaThrMetGlyLeu	1152
Db	2531	CTTGCTGTGCTTCACTCAGAGCATGCGCAAGGCCCCATGAGGCTCTTCAACATGGAGCTG	2590
QY	1153	ProLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGlnLeu	1172
Db	2591	CCCTTGCCTGGACCTTAAGAAAGCTG-----	2617
QY	1173	SerProArgGlyGlnAlaGlyProProGlyLysSerLeuGlyValProThrAlaGlnGluAla	1192
Db	2617	-----	2617
QY	1193	SerValLeuArgGlyThrAlaLeuGlyLysSerValProGlyGlySerIleThrLysGlyIle	1212
Db	2618	-----GGGACGACGACTGGGCTCCGACACAGTGAAGCATCCAAAGGGCTTC	2665
QY	1213	ProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGly	1232
Db	2666	CCCAAGTACCCGGGCTCCAGCGGCCCC---AGCTTACAGAGGCTCTTATCCACCGCG	2719
QY	1233	ThrProAlaAspValLeuLysTyrLysGlyThrIleThrArgLysIleGlyGlnLysSerPro	1252
Db	2720	ACGCGCGCAGACGTCTCTCAAGAGGTATCCATCAGCAGGATGTGTGGTGAAGACAGCCCA	2779
QY	1253	SerArgLeuAspArgGlyLysArgLysSerLeuProLysGlyHisValIleTyrGlyGly	1272
Db	2780	AGTGCCTTTGACCGGGCAGAGAGCACAACCTGCCCAAGGCGCAATGTATCATATGAGGCG	2839
QY	1273	LysLysGlyHisValLeuSerLysTyrGlyGlyGlyMetSerValThrGlnLysSerLysGlu	1292
Db	2840	AAGAAAGGCCACGTCTTATCTTATGAAGGTGTATGTCTGGTGTACAGAGCTTAAAGAG	2899

QY 1293 AspglyArgSerSerSerGlyProProHisgluThrAlaAlaProValArgThrTyAsp 1312
DB 2900 GATGAGAAAGACAGCTCGAGCCACCACTGAGACTGCGCCCTTAACAGACCTATGAC 2959
QY 1313 MetMetGluGlyArgValAlaArgAlaIleSerSerAlaSerIleGluGlyLeuMetGly 1332
DB 2960 ATGATGAGAGGCGCGTGTAGGACGACTGACCTCAGCAGATGAGAGGACTCATGGCC 3019
QY 1333 ArgAlaIleProProGluArgHisSerProHisIleuLeuGluGlnHisIleArg 1352
DB 3020 CGCGCATC-----CCTGAGCAGACAGCCCC--CACCTCAAGAGACAGATCACTACCGA 3073
QY 1353 GlySerIleThrGlnGlyIleProArgSerTyValGluAlaGlnGluAspTyLeuArg 1372
DB 3074 GGCCTCATCAGCAAGGACATCCGAGGTCTATGTGAGAGCCAGAGAGACTATCTTACGG 3133
QY 1373 ArgGluAlaIleuLeuLeuArgGluGlyThrProProProProProSerArgAsp 1392
DB 3134 CGGAGGCCAAGCTCTTGAAGGAGAGAGGACACCAACCCCAACCACTCGGGAGC 3193
QY 1393 LeuThrGluAlaTyIleTyThrGln-----AlaLeuGlyProLeuLeuLeuLeuPro 1409
DB 3194 CTGACTGAGACTCAAGCCCGGCCCCCTGACCCCTGCGGTCCCTGAAAGCTGAAGCCG 3253
QY 1410 AlaHisgluGlyLeuValAlaThrValIlysgluAlaGlyArgSerIleHisgluIlePro 1429
DB 3254 ACTCAGAGGGGTGTGTAGCAACTGTGAAGAGGGGGCGGCTCATTCATAGATCCCG 3313
QY 1430 ArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLeuGluGly 1449
DB 3314 AGAGAGAGACTCGCCGCGACACTGAGCTACCCCTGGACCAACGGCTCTGAAGAGGGGT 3373
QY 1450 SerIleThrGlnGlyThrProLeuLeuTyIleAspThrGlyAlaSerThrThrGlySerIys 1469
DB 3374 TCCATCACCCAGGGGACCCCACTCAAGTACGACTCTGGGGACACCCCTCACTGGCAACAG 3433
QY 1470 IysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisPro 1489
DB 3434 AAACAGAGGTGGCTCCATCATCGGACGCCCCGGGGCGCTTCCCTGCGCACCCG 3493
QY 1490 LeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyIleGluIleSerLeuIys 1509
DB 3494 CTGGACATATAGCTGATGCCCGGGACCTGGAGCGTCCGCTATGAAAGAGTCTGAAG 3553
QY 1510 SerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIle 1529
DB 3554 ACCCGGTCAAGGACGACAGTGTGTGAGGGGGCTCCATCAACAGTGGGGTCCAGTCTC 3613
QY 1530 ValProGluLeuGlyIysProArgGlnSerProLeuThrTyIleGluAspHisgluAlaPro 1549
DB 3614 GTGCTGAACTGGGCAAGCCAGCAAGCCACTGACTTACAAAGACACGGGGGACCC 3673
QY 1550 PheAlaGlyHisLeuProArgGlySerProValIleMetArgGluProThrProArgLeu 1569
DB 3674 TTCACACAGTCACTGCGACAGTGGCTCCCTGTGACACAGAGGAGCCAGCACCGCTT 3733
QY 1570 GlnGluGlySerLeuSerSerSerIysAlaSerGlnAspArgIleuLeuThrSerThrPro 1589
DB 3734 CAGGAAGGACGCTCTCATCAAGCAAGCGCTCCAGGACCGGAAGCTGACATCTACACCC 3793
QY 1590 ArgGluIleAlaIysSerProHisSerThrValProGluHisIleProHisProIleSer 1609
DB 3794 CGGGAATTCGCCAAGTCCCAACAGACATGTGCGCAGACACCACTTCAACCCCATCTCC 3853
QY 1610 ProTyIleGluHisLeuLeuArgGlyValSerGlyValAspLeuTyIleArgSerHisIlePro 1629
DB 3854 CCTTATGAGACTTCTCCCGGGCGTGTACGTGTGACCTGTATCCGTGTGCATATCCA 3913
QY 1630 LeuAlaPheAspProThrSerIleProArgGlyIleProLeuAsp---AlaAlaAlaAla 1648
DB 3914 TTGGCCCTTTGACCCCACTCATACCCGAGGAGTCCCTCTGGAAAGACAGCGCGCAGCC 3973
QY 1649 TyIleTyIleuProArgHisLeuAlaProAspProThrTyIleProHisLeuTyIleProProTyIle 1668

DB 3974 TACTACCTGCCCCGGGACCTTGCCCCCGACCCCACTTACCAACACTGTATCCACTTAC 4033
QY 1669 LeuIleArgGlyTyIleProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAsp 1688
DB 4034 CTCAATCGGGCTACCTCTGACACGGGGCGCTTGAGAAACCCCAACCATCATCATATAC 4093
QY 1689 TyIleThrSerGlnGlnMetHisIleAsnThrAlaThrAlaMetAlaGlnArgAlaAsp 1708
DB 4094 TACATCATCTCGACAGACATGACACCAACGCTGCTCCGACATGGCCACAGCTGCTGAC 4153
QY 1709 MetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyIleAlaGlyPro 1728
DB 4154 ATGCTGAGGGGTCTGTCAACCCGAGATCTCCGCTGGCCCTCAATATGCGCTGGCCCA 4213
QY 1729 ArgGlyIleIleAspLeuSerGlnValProHisLeuProValIleValIleProProThrPro 1748
DB 4214 AGAGGCAATTATCGACTTCTCCAAAGTGCACACCTGCCGTGTGTGTGCACCAACGCCA 4273
QY 1749 GlyThrProAlaThrAlaMetAspArgLeuAlaTyIleuProThrAlaProGlnProPhe 1768
DB 4274 GGACACCTTGCCACCGGCATGACCGGCTTGCTTACTCTCCACTGGCCCCCACTTCC 4333
QY 1769 SerSerArgHisSerSerSerProLeuSerProGlyIleProThrHisIleuThrIysPro 1788
DB 4334 AGCAACCCCAACAGTACGCTCACCGCTGTCCAGAGAGGCCCACTACCTTAACCA 4393
QY 1789 ThrThrThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspArg 1808
DB 4394 ACTGCCACATCTTCAATCGAGGGAGAGGAGGAACTGAGCGGGAACGAGAC----- 4444
QY 1809 GluArgGluIysSerIleLeuThrSerThrThrThrValGluHisAlaProIleTyIleArg 1828
DB 4445 -----AGTCCATCTCAAGTCTACACTACATGATGACATGACCCATCTGGACA 4495
QY 1829 ProGlyThrGlnGlnSerSerGlySerSerGlySerGlyGlyGlyIleGlySerSer 1848
DB 4496 CTGTGTACGAGACAGACGACGGGGCT-----GGGGGCACG 4531
QY 1849 SerArgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGln 1868
DB 4532 AGCCGCCCCGCTCCACACC-----CACAGCACTGCCCATCTCCCGGGACCCAG 4585
QY 1869 AspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetIlyGlyIleIleThr 1888
DB 4586 GAGCGCTTGACACAGAGGCCAGTGTGTCACAACAGACATGAAGGGCTGTGTACCC 4645
QY 1889 Ala-ValGluProSerIysProThrValLeuArgSerThrSerThrSerProValArg 1908
DB 4646 TCCCGTGAACCCGGGACCCCAACGCTCCTGAGGTCCACCTCCACTTTCGCTGTCCG 4705
QY 1908 GProAlaAlaThrPheProProAlaThrHisCysPProLeuGlyGlyThrLeuAspGlyVal 1928
DB 4706 CCAGCTGCACATTCACCTGACCAACCCCACTGCCCACCTTGTGTGSCACCTTGAAGGGGT 4765
QY 1928 IlyProThrLeuMetGluProValLeuLeuProIlyGluAlaProArgValAlaArgPro 1948
DB 4766 CTACCTTACCTCATGAGACCGGTCTCTGTATCCAAAGAGACCTTCGGGTGCCCGGCC 4825
QY 1948 OGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaIysProProAlaArgSerGly 1968
DB 4826 CGAGCGGCCCGGTGTGAGCGTGGCCATGCTTCTCCACCAACCCCGCGCGG----- 4880
QY 1968 IlyLeuGluProAlaSerSerProSerIysGlySerGluProArgProLeuValProProVa 1988
DB 4881 ----GAGCCGCTCTCTCAACCAAGAAAGCTCGAGCCCCCATCTCTAGACACCCCCAG 4936
QY 1988 IserGlyHisAlaThrIleAlaArgThrProAlaIysAsnLeuAlaProHisIleAlaSe 2008
DB 4937 CTCACACCAACAGCATGCGCGGACCCCAAGAAAGACCTTGCACCCACCATGCGCAG 4996
QY 2008 rProAspProProAlaProProAlaSerAlaSerAspProHisIleArgGluTyThrGlnSe 2028

Db 4997 TCCGAGCCGCCGCG---CCCACTCGGCTCAGATCTGACCGAGAAAAGACTCAAG 5053
Qy 2028 rlysrProPheSer11eGlnGluLeuGluLeuArgSerLeuGlyTyrHis--GlySerSe 2047
Db 5054 TAAACCTTTTCCATCCAGAAATTGGAATCCGTTCTCTGGGTTTACACAGTGAAGCTCG 5113
Qy 2047 rTyrSerProGluGlyValGluProValSerProValSerSerProSerLeuThrHis 2067
Db 5114 CTACAGCCCCGATGAGGTGAGCCCATCAGCCCGGTGAGCTCCCGACCTGACCCACGA 5173
Qy 2067 pLySGlyLeuProLyHis1leuGluGluLeuAspLySerHis1leuGluGlyLeuLeu 2087
Db 5174 CAAGGGGCTCTCCAAACCTCTGGAAGAGCTGAAGAAAGCCACTTGAAGGGAGCTGCG 5233
Qy 2087 gProLyGlnProGlyProValLyLeuGlyGlyGluAlaHis1leuProHisLeuAr 2107
Db 5234 GCACAAAGCAGCCAGCCCATGAAAGCTCAGCGGGAGGCTGCCCATCTCCACATCTGCG 5293
Qy 2107 gProLeuProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyVal 2127
Db 5294 GCCACTGCCGAGAGCCAGCCCTCATCCAGCCCATCTCCGACTGCCCCAGGCACTCA 5353
Qy 2127 sGlyHisGlnArgValValThrLeuAlaGlnHis1leuSerGluVal1leuThrGln 2147
Db 5354 AGGTCAACAGAGGGTGGTCACTCTGGCTCAGCATCAGCAGAGGTCTTACGACGACTA 5413
Qy 2147 rThraGHis1leuProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPhe 2167
Db 5414 CAGCGCCACACCCGACAGACTCACTGCGCCCTCCGACCTCTCACTCTCTTCC 5473
Qy 2167 oGlyAlaSerCyAsProValLeuAspLeuArgArgProProSerAspLeuTyrLeuPro 2187
Db 5474 CCGAGCCAGCTGCTCTGCTGATCTTCCGCCGCCACCGACGACTTCACTCCACCC 5533
Qy 2187 oProAspHisGlyAlaProAlaArgGlySerProHisSerGlyGlyLyAspArgSer 2207
Db 5534 CCCCAGCAAGGAGCCAGCCCGGGATCCCCCAAGAAAGGGGGCAAAAGGTCCCC 5593
Qy 2207 oGluProAsnLyThrSerValLeuGlyGlyGlyGluAspGly1leuGluProValSer 2227
Db 5594 AGAACCCAGCAAAACATCGTCTCTGGGAGAGTGAAGATGCCATTTAGCTGTGTC 5653
Qy 2227 oProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyr 2247
Db 5654 ACCAGAGGGGATCATGAGCCAGACATGCTCGAAGGCTGTATCCCATCTGATATCG 5713
Qy 2247 gAspGlyGluGlnThrGluProSerArgMetGlySerLySerProGlyValAsnThr 2267
Db 5714 AGACGGGGAACAGGGGAGGCC---AGATGGGCTTAAGTCTCAAGGCAACACGACCA 5770
Qy 2267 nProProAlaPhePheSerLyLeuThrGlySerAsnSerAlaMetValLySerLy 2287
Db 5771 GCCGCCAGCTTTCTTCACTAGTGAAGTGAAGCACTCCGCAATGTTGAAGTGAAGAA 5830
Qy 2287 sGlnGlu1leuAsnLyLeuLeuAsnThrHisAsnArgAsnGluProGlyTyrAsn1le 2307
Db 5831 GCAGGAGATCAACAAACTCAACCCCAACCGGAAGACAGCAAAATATTTG 5890
Qy 2307 rGlnProGlyThrGlu1leuPheAsnMetProAla1leuThrGlyThrGlyLeuMet 2327
Db 5891 CCAGCTCGGAGCGGAATCTTCAACATGCCCCGACATCATCGAGCGGCTTATGAC 5950
Qy 2327 rArgSerGlnAlaValGlnGluHis1leuAsnThrAsnMetGlyLeuGluAla1leuLe 2347
Db 5951 TGAAGCCAGGCGCTGCAAGAACACGACACCAATAGGGCTGAAGGCCATTTATG 6010
Qy 2347 gLyVal1leuMetGlyLySerTyrAspGlnTrrGluGluSerProProLeuSerAlaAs 2367
Db 6011 AAAGGACATCATGCTGAATATATGATCATGTGGAAGGCCCGCCGCTCGGCGCCATG 6070
Qy 2367 aPheAsnProLeuAsnAlaSerAlaSerLeuPro--AlaAlaMetPro1leuThrAla 2386
Db 6071 TTTTAACCTCTGAATGCGAGCGCAGTGTGCGCTGTATGCCCCATTAACCACTGC 6130

Qy 2386 aAspGlyArgSerAspHis1leuThrSerProGlyGlyGlyLyVal1leuValSe 2406
Db 6131 TGAAGAGGAGTGAACCAACCACTCACTGCGAGAGTGAAGTGAAGAAAGGTCTC 6190
Qy 2406 rGlyArgProSerSerArgLyVal1leuSerProAlaProGlyLeuAlaSerGlyAs 2426
Db 6191 TGGCAGACCTAGACCGGAAGAAAGTGCAGACCAAGGCTTACCTCGGAGACCG 6250
Qy 2426 gProProSerValSerSerAlaHisSerGlyGlyAspCyAsnArgArgThrProLeu 2446
Db 6251 ACCCTCTGCTCTCTCACTACATCAAGGGGAGTGAATGCGGAACACCACTCAC 6310
Qy 2446 rAsnArgValTrrGlyAspArgProSerSerAlaGlySerThrProPheProTyrAs 2466
Db 6311 CAACCGTGTGGAGAGACCGGCCCTCATCTGAGAGGTCAACCCATCTCCCTCAACCC 6370
Qy 2466 oLeu1leuMetArgLeuGlnAlaGlyValMetAlaSerProProProGlyLeuPro 2486
Db 6371 TTTGATTATGAGGCTACAGGAGATGTCATGCTCCCGCCCACTGGCTTGCGGC 6430
Qy 2486 aGlySerGlyProLeuAlaGlyProHis1leuAsnAlaTrrAspGlyGluProLy 2506
Db 6431 AGGACAGGGGCCCTTAGCTGTGCCACCAAGCTGGAGTGAAGAGCCCAAGCACTGCT 6490
Qy 2506 uCySerGlnTyrGluThrLeuSerAspSerGlu 2517
Db 6491 GTCTTCAAGTATGAGACACTCTCGACAGCGAG 6524

RESULT 14
US-10-146-473-14
; Sequence 14, Application US/10146473
; Publication No. US2003010888A1
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Matthew
; APPLICANT: Gout, Ivan
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Gure, Ali
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Old, Lloyd
; TITLE OF INVENTION: Breast Cancer Antigens
; FILE REFERENCE: L00461/70130 (JRV)
; CURRENT APPLICATION NUMBER: US/10/146, 473
; CURRENT FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/291, 150
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 14
; LENGTH: 2930
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-146-473-14

Alignment Scores:
Pred. No.: 7, 66e-264 Length: 2930
Score: 4612.00 Matches: 873
Percent Similarity: 94.38% Conservative: 1
Best Local Similarity: 94.28% Mismatches: 6
Query Match: 34,90% Indels: 46
DB: 16 Gaps: 1

US-09-522-753-5 (1-2517) x US-10-146-473-14 (1-2930)

Qy 1592 lIeAlaLySerProHisSerThrValProGluHis1leuProHisProHisSerPro 1611
Db 123 CTGGCTCAACCCCGACAGCAAGCTGCGGAGCACACCAACCCCACTTCTGCGCTTAT 182
Qy 1612 GluHisLeuLeuArgGlyValSerGlyValaAspLeuTyrArgSerHis1leuProLeu 1631
Db 183 GAGCACCTCTCTGGGGCGTGAAGTGGAGTGAACCTGTATGCGAGCCACATCCCTCGGCC 242
Qy 1632 pHeAspProThrSer1leuProArgGly1leuProLeuAspAla1leuAlaTyrTyrLeu 1651

|||||
Db 243 TTCGACCCCACTCCATACCCCGCGGCATCCCTCTGAGACGACGCCCTCACTACCTG 302
Qy 1652 ProAaRgH1sLeuAlaProAsnProThrTyrrProH1sLeuTyrrProProTyrrLeuH1eAaRg 1671
Db 303 CCCCAGACCTGAGCCCCCAACCCCACTTACCCGACCTTGTAACCCACCTTACCTCACTCCG 362
Qy 1672 G1yTyrrProAspThrAlaAlaLeuGluAaAsnArgGlnThr11e11eAsnAspTyrr11eThr 1691
Db 363 GGCTACCCCGACACGCGCGGCGTGGAGAACCGGCACATCATCATCAATGACTACATCAC 422
Qy 1692 SerGlnGlnMetH1sH1sAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuAaRg 1711
Db 423 TCGACAGCATGACACACACGCGGCACCGCATGGCCACGAGACTGATATGCTGAAG 482
Qy 1712 G1yLeuSerProArgGluSerSerLeuAlaLeuAsnTyrrAlaAlaG1yProArgG1yLe 1731
Db 483 GGCTCTCGCCCCGACAGTCTTGCTGGACTCACTTACGCTGCGAGTCCCGAGGCATC 542
Qy 1732 11eAspLeuSerGlnValProH1sLeuProValLeuValProProThrProG1yThrPro 1751
Db 543 ATCGACCTGTCCCAATGACACACCTGCGTGTCTGTGCCCCGACACACGAGCACCCCA 602
Qy 1752 AlAThrAlaMetAspArgLeuAlaTyrrLeuProThrAlaProGlnProPheserSerArg 1771
Db 603 GCCACCGCATGGACCGCTTGCTACTTCCCAACCGCGCCCAAGCCCTTCAGACACCCGC 662
Qy 1772 H1sSerSerSerProLeuSerProG1yG1yProThrH1sLeuThrTyrrProThrThr 1791
Db 663 CACAGCAGCTCCCACTCTCCAGAGAGTCCAAACACTTGACAAACACACACACACAG 722
Qy 1792 SerSerSerG1yAaRgG1yAaRgAspArgAspArgG1yAaRgAspArgG1yAaRgG1y 1811
Db 723 TCCTGCTCCGAGGGAGGAGACCGGATCGAGACGGGACCGGATCGGAGCGGAG 782
Qy 1812 TyrrSer11eLeuThrSerThrThrThrValGluH1sAlaPro11eThrArgProG1yThr 1831
Db 783 AAGTCAATCTCACTGACACACAGAGGTGAGACAGCACCCATCTGAGACCTGTGTACA 842
Qy 1832 G1yGlnSerSerG1ySerSerG1ySerSerG1yG1yG1yG1ySerSerSerArgPro 1851
Db 843 GAGCAGACGAGCGGACGACGCGACAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 902
Qy 1852 AlAspSerH1sSerH1sAlaH1sGlnH1sSerPro11eSerProArgThrGlnAspAlaLeu 1871
Db 903 GCCTCCCACTCCATGCCACACAGCATGTGCCATCTCCCTCGACCCACAGATGCCCTC 962
Qy 1872 GlnGlnArgProSerValLeuH1sAsnThrG1yMetG1y11e11eThrAlaValGlu 1891
Db 963 CAGCAGAGACCCAGTGTGCTTCAACACAGGCATAGAGGTATCATCACCGCTGTGAG 1022
Qy 1892 ProSerLyProThrAlaLeuArgSerThrSerSerSerProAlaArgProAlaAla 1911
Db 1023 CCAAGACGCGCCAGGCTTGAAGTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCC 1082
Qy 1912 ThrPheProProAlaThrH1sCyrrProLeuG1yG1yThrLeuAspG1yValTyrrProThr 1931
Db 1083 ACATTCACACCTGACCCACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTTACC 1142
Qy 1932 LeuMetGluProValLeuLeuProLyG1yAlaProArgValAlaArgProG1yAaRgPro 1951
Db 1143 CTCATGAGACCGGCTTGTGCTTGCACCAAGAGAGGCCCGCGGTGCGCGGCGAGAGCGGCC 1202
Qy 1952 ArgAlaAspThrG1yH1sAlaPheLeuAlaTyrrProProAlaArgSerG1yLeuGlnPro 1971
Db 1203 CAGAGCAGACCGGCGCATGCTTCTGCGACACCCCGCGCTCGGAGTGGAGCCC 1262
Qy 1972 AlAspSerProSerLyG1ySerG1yProArgProLeuValProProValSerG1yH1s 1991
Db 1263 GCCTCTTCCCGACGAGGGCTCGAGGCCCGCGCTAGTGTCTCTGTCTGTGCGAC 1322
Qy 1992 AlAThr11eAlaArgThrProAlaTyrrAsnLeuAlaProH1sH1sAlaSerProAspPro 2011
|||||

Db 1323 GCCACCATCGCCCGACCCCTTCGAAAGAACTTCGCACTCACACGACCCGAGCCCG 1382
Qy 2012 ProAlaProProAlaSerAlaSerAspProH1sArgG1yTyrrThrGlnSerLySerProPhe 2031
Db 1383 CCGGCGCACCTGCTCGGCTCGGACCCGACCCGGAAGAAAGACATCAAGGTAAACCTTTT 1442
Qy 2032 Ser11eGlnGlnLeuGlnLeuArgSerLeuG1yTyrrH1sG1ySerSerTyrrSerProGlu 2051
Db 1443 TTCATTCAGGAACCTGAATCCGTTCTTGGGTTTACACGAGCAGACTACAGCCCGAA 1502
Qy 2052 G1yValGluProValSerProValSerSerProSerLeuThrH1sAspLyG1yLeuPro 2071
Db 1503 GGGGTGAGCCCGCTGACGCCCTTGAGCTTACCACTGTACCCACGACAAAGGGCTCCCC 1562
Qy 2072 TyrrH1sLeuGluGlnLeuAspLySerH1sLeuGlnG1yG1yLeuAaRgProLyGlnPro 2091
Db 1563 AAGCACCCTGAAAGAGTGCACAAAGGCCACTGAGAGGGGAGCTGCGGCCCAAGAGCCA 1622
Qy 2092 G1yProValTyrrLeuG1yG1yGluAlaAlaH1sLeuProH1sLeuAaRgProLeuProGlu 2111
Db 1623 GGCCCGTGAACTTGCGCGGAGGCGCCCACTTCCACACTGCGCGCGCTGCTGAG 1682
Qy 2112 SerGlnProSerSerSerProLeuLeuGlnThrAlaProG1yValLyG1yH1sGlnArg 2131
Db 1683 AGCCAGCCCTCGTCCAAGCCCGTGTCCAGACCCGCCAGGGGTCAAGGTCAACAGCGG 1742
Qy 2133 ValValThrLeuAlaGlnH1s11eSerGluVal11eThrGlnAspTyrrThrArgH1sH1s 2151
Db 1743 GTGTACACCTTGCCACAGCATCAAGTATCATCACACAGACTACACCCGGACAC 1802
Qy 2152 ProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrrSerPheProG1yAlaSerCyrr 2171
Db 1803 CCACAGCAGCTCAGGCGACCCCTGCGCCGCCCTTACTCTTCTCTGAGGCGCACTGC 1862
Qy 2172 ProValLeuAspLeuArgArgProProSerAspLeuTyrrLeuProProProAspH1sG1y 2191
Db 1863 CCGTCTCGAGCTTCGCGCGCCACCACTGATACCTTACTCTCCCTGAGGCGCACTGC 1922
Qy 2192 AlAProAlaArgG1ySerProH1sSerG1yG1yLyAaRgSerProGluProAsnLy 2211
Db 1923 GCCCGGCGCGTGGCTCCCGCCACAGCGAAAGGGGAGAGGTCTCCAGAGCCAAACAAG 1982
Qy 2212 ThrSerValLeuG1yG1yG1yGluAspG1y11eGluProValSerProProGluG1yMet 2231
Db 1983 AGTGGCTTGTGGGTGTGTAGGACGGTATTTGAACCTGTCTCCACCGGAGGGCAAG 2042
Qy 2232 ThrGluProG1yH1sSerArgSerAlaValTyrrProLeuLeuTyrrArgAspG1yGluGln 2251
Db 2043 AGGAGGCCAGGCGACTCCCGAGTGTGTATCCGCTGTACCGGATGGGAGACAG 2102
Qy 2252 ThrGluProSerArgMetG1ySerLySerProG1yAsnThrSerGlnProProAlaPhe 2271
Db 2103 ACGAGGCCACAGAGATGGGCTCCAAGTCTCAAGGAAACACAGCAGCGCGCAGCTTC 2162
Qy 2272 PheSerLyLeuThrGluSerAsnSerAlaMetValTyrrSerLyGlnGln11eAsn 2291
Db 2163 TTCAGAGACTGACCGAGAGCAATCCCGCAATGATCAATCCAAAGAGCAAGATCAAC 2222
Qy 2292 TyrrLyLeuAspThrH1sAsnArgAsnGluProGluTyrrAsn11eSerGlnProG1yThr 2311
Db 2223 AAGAAAGCTGAACCCACCAACCGGAATGAGCTGAATACAAATTCAGCCAGCTGGAGAG 2282
Qy 2312 G1y11ePheAsnMetProAla11eThrG1yThrG1yLeuMetThrTyrrArgSerGlnAla 2331
Db 2283 GAGATCTTCAATATGCGCGCATCAACCGGAACAGGCTTATGACTATGAAAGCCAGGCG 2342
Qy 2332 ValGlnGlnH1sAlaSerThrAsnMetG1yLeuGluAla11e11eArgLyValAlaLeuMet 2351
Db 2343 GTGACGAGAAATGACAGACCAACATGGGGCTGAGGCCATATTAAGAAAGCACTCAAG 2402
Qy 2352 G1yTyrrTyrrAspGlnThrProGlu11eSerProProLeuSerAlaAsnAlaPheAsnProLeu 2371
|||||
Db 2402 ----- 2402

QY 2372 AsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAsp 2391
Db 2402 ----- 2402
QY 2392 HisThrLeuThrSerProGlyGlyGlyValValAlaValSerGlyArgProSerSer 2411
Db 2403 -----GGTGGCGGCGGAGGCCCAAGGCTCTGGCAGACCAGCAGC 2444
QY 2412 ArgValAlaValSerProAlaProGlyLeuAlaSerGlyValAspArgProProSerValSer 2431
Db 2445 CGAAAGCCCAAGTCCCGGCGCCGCGCTGGCATCTGGAGCCGCGCACCTCTGTCTCC 2504
QY 2432 SerValHisSerGlyGlyAspCysAsnArgArgThrProLeuThrAsnArgValTyrGlu 2451
Db 2505 TCAGTGCACTCGAGGGAGAGACTGCACCGCGGAGCGCGCTCACCAACCGGCTGGAG 2564
QY 2452 AspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeu 2471
Db 2565 GACAGGCGCTCGTCCGAGGTTCCAGCGCCATTCCTTACCAACCCCTGATCATCGGCTG 2624
QY 2472 GluAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeu 2491
Db 2625 CAGGCGGCTCATGGCTTCCACCCCGGCGCTCCCGGCGGAGCGGCGCTC 2684
QY 2492 AlaGlyProHisAlaValTyrAspGlyGlyProArgProLeuLeuCysSerGlnTyrGlu 2511
Db 2685 GCTGGGCGCCACCGCTCGGAGAGAGAGCCCAAGCCATGCTCTCTGGCACTGAG 2744
QY 2512 ThrLeuSerAspSerGlu 2517
Db 2745 ACACTCTCCGACAGCGAG 2762

RESULT 15

US-10-174-014-13
Sequence 13, Application US/10174014
Publication No. US20040005292A1
GENERAL INFORMATION:
APPLICANT: C. Frank Bennett
APPLICANT: Susan M. Freiler
APPLICANT: Kenneth W. Dobie
TITLE OF INVENTION: ANTISENSE MODULATION OF SMRT EXPRESSION
FILE REFERENCE: PFS-0012
CURRENT APPLICATION NUMBER: US/10/174, 014
CURRENT FILING DATE: 2002-06-17
NUMBER OF SEQ ID NOS: 73
SEQ ID NO 13
LENGTH: 2930
TYPE: DNA
ORGANISM: H. sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (456) ... (2765)
US-10-174-014-13

Alignment Scores:

Pred. No.: 7,66e-264 Length: 2930
Score: 4612.00 Matches: 873
Percent Similarity: 94.38% Conservative: 1
Best Local Similarity: 94.28% Mismatches: 6
Query Match: 34,90% Indels: 46
DB: 18 Gaps: 1

US-09-522-753-5 (1-2517) x US-10-174-014-13 (1-2930)

QY 1592 IleAlaValSerProHisSerThrValProGluHisHisProHisProIleSerProTyr 1611
Db 123 CTGGGCTCACCCCGGACAGACACGTCGCCGAGACACCAACCCCATCTGCCTAT 182
QY 1612 GluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAla 1631
Db 183 GAGCACTGCTTGGGCGGTCGAGTGGCGTGCATCTGATGCAAGCCATCCCTGGCC 242

QY 1632 PheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeu 1651
Db 243 TTGACCCCACTCCATACCCCGGCGATCCCTCTGAGCCAGCGCTGCTACTACTG 302
QY 1652 ProArgHisLeuAlaProAsnProThrTyrProHisIleuTyrProProTyrLeuIleArg 1671
Db 303 CCCGACACCTGGACCCCAACCCCACTTACCCGACCTGTATACCACTTACTATCCGC 362
QY 1672 GlyTyrProAspThrAlaAlaIleuGluAsnArgGlnThrIleIleAsnArgTyrIleThr 1691
Db 363 GGTACCCCAACCGGCGGCTGGAGAACCGGAGCCATCATCATGACTATATACC 422
QY 1692 SerGlnIleMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArg 1711
Db 423 TCCGAGAGATGACACCAACCGGCGCATGGCCATGGCCAGAGCTGATATGCTAGG 482
QY 1712 GlyLeuSerProArgGlySerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIle 1731
Db 483 GGCCTCTGCGCCGCGAGTCTCGTGGCACTCACTACGCTGGGGTCCCGAGGATC 542
QY 1732 IleAspLeuSerGlnValProHisIleuProValLeuValProProThrProGlyThrPro 1751
Db 543 ATGACCTGTCCAGAGTGCACCTGCTGTGCTGTCGCCCGACACAGGACCCCA 602
QY 1752 AlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerArg 1771
Db 603 GCCACCGCATGAGACCGCTTGCCTTACTCCACCGCGCCCATTCAGACGCGC 662
QY 1772 HisSerSerSerProLeuSerProGlyGlyProThrHisIleuThrTyrProThrThr 1791
Db 663 CACAGAGAGTCCCACTCTCCAGAGGTCACACACTTGAACAAACCAACACAG 722
QY 1792 SerSerSerGluArgGluArgAspArgGluArgAspArgAspArgGluArgGlu 1811
Db 723 TCCTCTCCGAGGGAGGAGCAGACCGGATCGAGAGGGAGCCGGATTCGGAGCGGAA 782
QY 1812 LysSerIleuThrSerThrThrThrValGluHisAlaProIleTyrAspProGlyThr 1831
Db 783 AAGTCATCTTCACGTCCACACGAGTGGAGACGACCACTTGGAACTTGATACA 842
QY 1832 GluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgPro 1851
Db 843 GAGCAGAGCAGCGGAGCAGACCGGACAGCGGGGGTGGGGGAGCAGACGCGGCC 902
QY 1852 AlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeu 1871
Db 903 GCCTCCCACTCCATGCCCCACAGCACTGCCATCTCCCTCGGAGCCAGATGCCCTC 962
QY 1872 GluGlnArgProSerValLeuHisAsnThrGlyMetLeuGlyIleIleThrAlaValGlu 1891
Db 963 CAGCAGAGACCAAGTGTGCTTCAACACAGGATTAAGGATATCATCCGCTGGAG 1022
QY 1892 ProSerLysProThrValLeuArgSerThrSerSerProValArgProAlaAla 1911
Db 1023 CCCAGACCGCCACGCTCGAGGTCCACCTCCACCTGTCACCGTGGCCAGGCGCC 1082
QY 1912 ThrPheProProAlaThrHisCysProLeuGlyIleThrLeuAspGlyValTyrProThr 1931
Db 1083 ACATTTCCCACTGCGCACCACTGCGGAGGAGACCTTCATATGGGTGCTTACCTTACC 1142
QY 1932 LeuMetGlnProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgPro 1951
Db 1143 CTATGAGCGCGCTCTTGCTGCCCAAGAGGCCCGCGGAGTCCCGGCGAGAGGCC 1202
QY 1952 ArgAlaAspThrGlyHisAlaPheLeuAlaValProProAlaArgSerGlyLeuGluPro 1971
Db 1203 CGAGCAGACACCGGCGCATGCTTCTCGGCAAGCCCGGCGGCTCGGAGGCC 1262
QY 1972 AlaSerSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHis 1991
Db 1263 GCTCTCTCCCAAGAGGCTCGAGGCCCGGCGCTTGTGTGCTCTGCTGCGAC 1322

QY 1992 AlaThrTLeuAlaArgThrProAlaIuysAanLeuAlaProHisAlaSerProAspPro 2011
DB 1323 GCCACCATCGCCCGGACCCCTCGAAGAACCTCGACCTCACACCCGACCCGCGACCG 1382
QY 2012 ProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPhe 2031
DB 1383 CCGGGCGCACCTGCTCGGCTCGGACCGGACCGGAAAAAGACTCAAGGTAAACCTTT 1442
QY 2032 SerTLeuGlnLysLeuGluLeuArgSerLeuGluTyrHisGlySerSerTyrSerProGlu 2051
DB 1443 TTCATCCAGGAACCTGAACCTCCGTTCTCTGGGTACACGACGACGACGACCCCGAA 1502
QY 2052 GlyAlaGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuPro 2071
DB 1503 GGGGTGACCCCGTCAGCCCTGTGAGCTCACCGAGCTGACCCACGACGAAGGGGCTCCCC 1562
QY 2072 LysHisLeuGluGlnLeuAspLysSerHisLeuGluGlnLysLeuArgProLysGlnPro 2091
DB 1563 AAGCACTGGAAGAGCTGCACAAAGACCACTGGAAGGGAGCTGCGGCCAAGACGCA 1622
QY 2092 GlyProValLysLeuGluLysGluAlaAlaHisLeuProHisLeuArgProLeuProGlu 2111
DB 1623 GGGCCCGTGAAGCTTGGCGGGAGGCGCCGACCTCCACACCTGGCGCGCTGCTGAG 1682
QY 2112 SerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGluHisGlnArg 2131
DB 1683 ACCCAGCCCTCGTCCAGCCGCTGCTCCAGACCGGCCCAAGGGGTCAAGGTCAACGCGG 1742
QY 2132 ValAlaThrLeuAlaGlnHisLeuSerGluValIleThrGlnAspLysThrArgHisSHis 2151
DB 1743 GTGGTACCTCGGCGCCGACATCACTGAGGTATACACAGGACTACACCCGCGACCAAC 1802
QY 2152 ProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGluAlaSerCys 2171
DB 1803 CCACAGCACTCAGCGCAACCCCTGCGCCGCCCTCTACTCTCTCTGCGGCGCACTGC 1862
QY 2172 ProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProAspHisGly 2191
DB 1863 CCGGTCTGGAACCTCGCGCGCCACCAAGTGAACCTTACTCTCGCGCGCGACATGGT 1922
QY 2192 AlaProAlaArgLysSerProHisSerGluGluLysArgSerProGluProAsnLys 2211
DB 1923 GCGCCGCGCGGTGGCTCCCGCACAGCGAAGGGGCAAGAGGTCTCCAGAGCCAAACAAAG 1982
QY 2212 ThrSerValLeuGluLysGluAspGlyLysGluProValSerProProGluGlyMet 2231
DB 1983 AGCTCGGTCTTGGGTGTGTGAGGACGGATTGAACCTGTGCCACCGAGGGCATG 2042
QY 2232 ThrGluProGluHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGln 2251
DB 2043 ACGGAGCCAGGGGCACTCCGAGTGTGTGACCCGCTGTACCGGGATGGGGAACAG 2102
QY 2252 ThrGluProSerArgMetGlySerLysSerProGluAsnThrSerGlnProProAlaPhe 2271
DB 2103 ACGGAGCCCAAGAGATGGGCTCAAGTCTCCAGGCAACCAAGCCAGCGCGCTTC 2162
QY 2272 PheSerLysLeuThrGlnSerAsnSerAlaMetValLysSerLysLysGlnLysLeuAsn 2291
DB 2163 TTCAGCAAGCTGACCGAAGCAACTCCGCAATGTCAAGTCCAAAGAACGAAGATCAAC 2222
QY 2292 LysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnLysSerGlnProGlyThr 2311
DB 2223 AAGAAGCTGAACCCCAACCGGAATGAGCTGAATATCAATATCAAGCCAGCTGGAGCG 2282
QY 2312 GluLysPheAsnMetProAlaIleThrGlyThrGlyLeuMetPheTyrArgSerGlnAla 2331
DB 2283 GAGATCTTCAATATGCGCGCATCACCGGAACAGGCTTATGACTATGAAGCCAGGCG 2342
QY 2332 ValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleLeuArgLysAlaLeuMet 2351
DB 2343 GTGCAAGAACATGCCAGCAACATGGGGCTGGAAGCCATAATTAGAAAGGCACTCATG 2402
QY 2352 GlyLysTyrAspGlnTyrGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeu 2371

DB 2402 ----- 2402
QY 2372 AsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAsp 2391
DB 2402 ----- 2402
QY 2392 HisThrLeuThrSerProGluLysGluGlyLysAlaLysValSerGlyArgProSerSer 2411
DB 2403 -----GTTGGCGCGGGAAGGCGCAAGGTCTCTGGCAGACCCGACAG 2444
QY 2412 ArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSer 2431
DB 2445 CGAAAAGCCAACTCCCGGCGCGGCTGSCATCTGGGGAACGAGCACCTCTGTCTCC 2504
QY 2432 SerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTyrGlu 2451
DB 2505 TCAGTGCACCTCGGAGGAGACCTGCACCGCGGAGCGCGCTCACCAACCGCGTGGGAG 2564
QY 2452 AspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuLysMetArgLeu 2471
DB 2565 GACAGGCCCTCGTCCGAGGTTCCAGGCAATTCCTTCAACCCCTGATCATGGCGCTG 2624
QY 2472 GlnAlaGlyValMetAlaSerProProProProGluLeuProAlaGlySerGlyProLeu 2491
DB 2625 CAGGCGGGGTGATGAGCTTCCACCCCAACCGGAGCTTCCCGGCGAGCGGCGCCCTC 2684
QY 2492 AlaGlyProHisHisAlaTyrAspGluGluProLysProLeuLeuCysSerGlnTyrGlu 2511
DB 2685 GCTGGCGGCCCAACAGCGCTGGGAGAGAGGCCAAGCCACTGCTGCTGCGAGTACGAG 2744
QY 2512 ThrLeuSerAspSerGlu 2517
DB 2745 ACACCTCCGACAGCGAG 2762

Search completed: November 12, 2005, 05:30:12
Job time : 6402 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 11, 2005, 09:36:02 ; Search time 1201 Seconds
(without alignments)
3429.233 Million cell updates/sec

Title: US-09-522-753-5
Perfect score: 13215
Sequence: 1 MSGSTQLVAVQWTRATEPRYP.....WDEPRKPLCSQYETLSDSE 2517

Scoring table:
BLOSUM62
Xgapop 10.0 , Ygapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues
Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+.p2n.model -DEV=x1p
-Q=/cgn2_1/USFTO.spool_p/US09522753/runat_02112005_173641_20819/app_query.fasta_1.2695
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFM=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09522753 @CNC 1.186 @runat_02112005_173641_20819 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCAL=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:*
1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/1/ina/6C_COMB.seq:*
6: /cgn2_6/prodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	12922	97.8	9053	4 US-09-976-594-306	Sequence 306, App
2	4165.5	31.5	7912	4 US-09-949-016-3804	Sequence 3804, Ap
3	726.5	5.5	33529	2 US-09-144-085-3	Sequence 3, Appli
4	718	5.4	1922	2 US-08-372-652-9	Sequence 9, Appli
5	718	5.4	1922	2 US-08-372-652-9	Sequence 9, Appli
6	698	5.3	14462	4 US-09-902-540-1090	Sequence 1090, Ap
7	698	5.3	14462	4 US-09-902-540-1090	Sequence 1090, Ap
8	689	5.2	4411529	3 US-09-103-840A-1	Sequence 1, Appli
9	685	5.2	4403765	3 US-09-103-840A-2	Sequence 2, Appli
10	670	5.1	44377	2 US-08-804-227C-7	Sequence 7, Appli
11	670	5.1	44377	2 US-08-804-227C-7	Sequence 1, Appli
12	668.5	5.1	4403765	3 US-09-103-840A-2	Sequence 2, Appli

13	665	5.0	4411529	3 US-09-103-840A-1	Sequence 1, Appli
14	661.5	5.0	43280	2 US-08-804-227C-1	Sequence 1, Appli
15	652	4.9	13987	2 US-08-804-227C-13	Sequence 13, Appli
16	648.5	4.9	50937	3 US-09-428-517-1	Sequence 1, Appli
17	641	4.9	28958	1 US-08-258-261B-6	Sequence 6, Appli
18	641	4.9	28958	1 US-08-258-261B-6	Sequence 6, Appli
19	641	4.9	28958	1 US-08-457-342-6	Sequence 6, Appli
20	641	4.9	28958	1 US-08-457-342-6	Sequence 6, Appli
21	641	4.9	28958	1 US-08-457-646A-6	Sequence 6, Appli
22	641	4.9	28958	1 US-08-458-076A-6	Sequence 6, Appli
23	641	4.9	28958	1 US-08-764-233A-4	Sequence 4, Appli
24	641	4.9	28958	1 US-08-457-335A-6	Sequence 6, Appli
25	641	4.9	28958	1 US-08-729-214-6	Sequence 6, Appli
26	641	4.9	28958	1 US-09-028-934-6	Sequence 6, Appli
27	631.5	4.8	49377	3 US-08-764-233A-1	Sequence 1, Appli
28	631.5	4.8	80161	3 US-09-370-200-1	Sequence 1, Appli
29	631.5	4.8	80161	3 US-09-370-200-1	Sequence 1, Appli
30	624.5	4.7	80161	3 US-09-603-207-1	Sequence 1, Appli
31	624.5	4.7	68750	3 US-09-335-409-1	Sequence 1, Appli
32	624.5	4.7	68750	3 US-09-568-102-1	Sequence 1, Appli
33	624.5	4.7	68750	3 US-09-567-969-1	Sequence 1, Appli
34	624.5	4.7	68750	3 US-09-568-480-1	Sequence 1, Appli
35	624.5	4.7	68750	3 US-09-568-486-1	Sequence 1, Appli
36	624.5	4.7	68750	3 US-09-568-472-1	Sequence 1, Appli
37	622.5	4.7	68750	3 US-09-567-899-1	Sequence 1, Appli
38	622.5	4.7	32241	4 US-09-502-540-5004	Sequence 5004, Ap
39	622.5	4.7	32241	4 US-09-502-540-1247	Sequence 1247, Ap
40	619	4.7	71989	3 US-09-443-501A-2	Sequence 2, Appli
41	616.5	4.7	47981	4 US-09-679-279-1	Sequence 1, Appli
42	614.5	4.7	8438	1 US-07-945-283-1	Sequence 1, Appli
43	614.5	4.7	15872	3 US-09-105-537-1	Sequence 1, Appli
44	614.5	4.7	15872	4 US-09-091-609-1	Sequence 1, Appli
45	609.5	4.6	17622	4 US-09-902-540-1125	Sequence 1125, Ap

ALIGNMENTS

RESULT 1
US-09-976-594-306
Sequence 306, Application US/09976594
Patent No. 6673549
GENERAL INFORMATION:
APPLICANT: Furness, Michael
TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
FILE REFERENCE: PA-0041 US
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: 60/240,409
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 1143
SOFTWARE: PERL Program
SEQ ID NO 306
LENGTH: 9053
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc:feature
OTHER INFORMATION: Incyte ID No. 6673549 898877.6
NAME/KEY: unsure
LOCATION: 2006, 2012
OTHER INFORMATION: a, t, c, g, or other
US-09-976-594-306
Alignment Scores:
Pred. No.: 0
Score: 12922.00
Percent Similarity: 98.26%
Best Local Similarity: 98.18%
Query Match: 97.78%
Length: 9053
Matches: 2487
Conservative: 2
Mismatches: 10
Indels: 35
Gaps: 5
DB: US-09-522-753-5 (1-2517) x US-09-976-594-306 (1-9053)

QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro 20
DB 520 ATGTGGAGATCCACACAGCGCTGTGGACAGAGCTGGAGGGCCACTGAGCCCGCTACCGG 579
QY 21 ProHlaSerLeuSerTyrTrpValGlnIleAlaArgThrHisGlnTrpValGlyLeuLeu 40
DB 580 CCCCAAGGCTTCTTCCATCCAGTGCAGATCGCCGAGACACACGAGCTGGGCTCTCCG 639
QY 41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisGluSerProGlySerIleIleGln 60
DB 640 GAGTACACAGACACTCCCGCACTATGCTCCCACTGTGCGCGGCTCCATCATCCAG 699
QY 61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
DB 700 CCCAGCGCGGAGGCGCTCCCTGTCTGTGAGTTCCAGCCGGGAATGAAACGATCCGAC 759
QY 81 GluLeuHisGluLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet 100
DB 760 GAGCTCCACCTGGCGGCAAGTCCCACTATACCTGCCGAGCTGGGGAAGTCCAGATG 819
QY 101 GluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
DB 820 GAGTTCATTGAAAGCAAGCGCCCTCGGCTAAGACTGCTGCTGACCCCTGCTGCCACCG 879
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
DB 880 TCACCCCTGTGGCCACGGGCGAGCTGCGGGATCTGAAGACTTCACCAAGACCGTAAAC 939
QY 141 LeuThrGlyLysLeuGluProValSerProProSerProProHisGlnTrpAspProGluLeu 160
DB 940 CTGACCGGGAAGCTGGACCGGTGTCTCCCGGACCGCCCGCACCTGAGCTTGAGCTG 999
QY 161 GluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAsp 180
DB 1000 GAGCTGTGCTCCGCAAGGCTGTCCAAAGAGAGACTATCCAGAACATGGAACCGCTGGAC 1059
QY 181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeu 200
DB 1060 CAGAGATACACCATGGTAGAGCAGCATCTTAACCTGAAGAAAGACGACAGCAGCTG 1119
QY 201 GluGluGluAlaAlaLysProProGluProGluLysProValSerProProIleGlu 220
DB 1120 GAGGAGAGAGCTGCCAAGCGCCGAGCTGAAGAACCCGATGCACCGCGCCCATCGAG 1179
QY 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAla 240
DB 1180 TCGAAGCACCGCAGCTGTGTGAGATCATCTACGACGAAACCGGAAGAGCTGAAGCT 1239
QY 241 AlaHisArgIleLeuGluGluLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPro 260
DB 1240 GCACATCGGATCTGGAAAGCTCGGGGCCCAAGTGAAGTGCCTGTGAACAACGAGCC 1299
QY 261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
DB 1300 TCCGACACCCGGGAGATCATGAGAACATCAAAATAAACGAGCGATGCGGAAGAGCTA 1359
QY 281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGln 300
DB 1360 ATCTTCTACTTCAGAGAGAGAAATCACGCTCGGAAACAATGGAGAGCAAGTTCTCCAG 1419
QY 301 ArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnPro 320
DB 1420 CCTATGACAGCTCATGAGGCGCTGGAGAAAGAAAGTGAAGCGCATTCGGAACAACCCC 1479
QY 321 ArgArgArgArgAlaLysGluSerLysValaArgLysTyrGluLysGlnPheProGluIle 340
DB 1480 CGCGGCGCGGCGCAAGAGAGCAAGTGCAGACTACTACAGAAAGAGATTCCTGAGATC 1539
QY 341 ArgLysGlnArgGluLeuGlnGluLysMetGlnSerArgValGlyGlnArgLysSerGly 360
DB 1540 CGCAAGCAGCGGAGCTGCAGAGGCGCATGCAAGCAGGGTGGGCGAAGCGGGCAATGGG 1599

QY 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
DB 1600 CTGTCCATGTCCGCGCGCCGACGACGAGCGGTGTGAGATCATCATGAGCCCTCTCA 1659
QY 381 GluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400
DB 1660 GAGCAGAGAACCGGAAGAGCAGATGCCAGCTGGCCGTATGCCGCCATGCTGTAC 1719
QY 401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420
DB 1720 GACGCTGACACAGCGCATCAAGTTTCATCACTGAATGAAGCGGCTATGGCCGACCCATG 1779
QY 421 LysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGlnTrpPhe 440
DB 1780 AAGGTGTACAAAGACCGCCAGGTCAATGAACATGTGAGAGCAGAGAGAAAGAACCTTC 1839
QY 441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
DB 1840 CGGAGAAAGTTCAATGACAGATCCCAAGAACTTTGGCTGATGCATATCTCTGAGAGAG 1899
QY 461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480
DB 1900 AAGACAGTGTGAGTGGGCTCTTATTACTACCTGACTGAAGAAGAACTAAG 1959
QY 481 SerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGln---GlnGlnGln 499
DB 1960 AACCTGTGTAGACGAGACTATCGCGCGCGCAAGAGCAGACAGACNACAGCAGCAG 2019
QY 500 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGln 519
DB 2020 CAGCAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2079
QY 520 LysAspGluLysGluLysGluLysGluAlaGluLysGluGluLysProGluValGlu 539
DB 2080 AAAAGATGAGAGAGAGAAAGAGGCGGAGAGAGAGAGAGAGAGAGCGGAGTGGAG 2139
QY 540 AsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGluLysAspAsp 559
DB 2140 AAGCAACAGAGAGACTCTCTCAAGAGAGAGACAGACACTCAGGGGAGGACCAACGAC 2199
QY 560 GluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLys 579
DB 2200 GAGAAAGAGCTGTGGCTCCAAAGCCGCAAAACCTGCCAAGCAGGGAAGACCCAAA 2259
QY 580 GlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGln 599
DB 2260 GCGCCGATCACCCGCTCAATGGCTAATGAGGCCAACAGGAGAGGCGCATCACCCCCAG 2319
QY 600 GlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpArgGluGlu 619
DB 2320 CAGAGCGCGAGACTGGCTCCATGAGCTGAATGAGAGTTCTCGCTGACAGAAAGAA 2379
QY 620 MetGluThrAlaLysLysGlyLeuLeuGlnHisGlyArgAsnTrpSerAlaIleAlaArg 639
DB 2380 ATGGAACAGCCAAAGAAAGTCTCTTGAAACACGCGCCGAACTGGTCCGCCATCGCCCG 2439
QY 640 MetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArg 659
DB 2440 ATGTGTGGCTCCAAAGCTGTGTGCGAGTGTAAAGAACTTCACTCAATCAAGAAAGAG 2499
QY 660 GlnAsnLeuAspGluIleLeuGlnGlnHisLysLysLeuLysMetGluLysGluArgAsnAla 679
DB 2500 CAGAACCTCGATGAGATCTTTCAGACGACCAAGCTGAAATGAGAAAGAGAGAAACGG 2559
QY 680 ArgArgLysLysLysLysAlaProAlaAlaAla--SerGluGluAlaAlaPheProProVal 699
DB 2560 CGAGAGAGAAAGAAAGAGCGCGCGCGCCAGCAGAGAGCTGCATTCGCCGCCGT 2619
QY 699 ValGluAspGluGluMetGluAlaSerGlyLysGlnGluGluLysGluMetValGly 719
DB 2620 GGTGTGAGATGAGAGATGAGAGCGGTGCGGCGTGAAGCAAAATGAGAGAGAGATGTGGA 2679
QY 719 uGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyTrp 739

Db 2680 GAGAGCTGAA----- 2689
Qy 739 oAlaThrValAsnAsnSerSerAspThrGluSerIleProSerProH1sthrGluAlaI 759
Db 2690 -GCCACTGTCAACACAGCTCAGACCCAGAGAGATCCCTCTCTCTCACTAGAGGCCG 2748
Qy 759 alyAspThrGlyGlnAsnGlyProlyProProAlaThrLeuGlyAlaAsp-GlyProP 779
Db 2749 CAAGGACACAGAGGAGATGGGCCCAAGCCCCAGCCACTCTGGGGCCGACGGGGCCAC 2808
Qy 779 rOProGlyProProProThrProPro-ArgArgThSerArgAlaProIleGlyProThrPro 798
Db 2809 CCCAGAGGCCACCCACCCACACCGAGAGCATCCGGGCC-CCCACTGAGCCACCCCG 2867
Qy 799 AlAserGluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaPro 818
Db 2868 GCGCTTGAGGCCACGGAGGCCCTAGCGCCCAACAGACCCCATCGCTCTGCACT 2927
Qy 819 ProProValValProlySGluGluGlyGluGlyGluThraAlaAlaProProValGlu 838
Db 2928 CTCTCTGTGGTCCCAAG 2987
Qy 839 GluGlyGlyGluGluGlyProProAlaAlaGluGluLeuAlaValAspThrGlyLyAsAla 858
Db 2988 GAGGGGGAG 3047
Qy 859 GluGluProValLySerSerGlyGlyThrGluGluGluGluGluGluGluGluGluGlu 878
Db 3048 GAGAGAGCCCTCAAGAGCGAGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3107
Qy 879 AspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuAlaGluGlyGluGlyGluGly 898
Db 3108 GAGCGAGAGGCCGTGAGGCCACCGCGAGGGGGCGCTCAAGAGAGAGAGAGAGAGAG 3167
Qy 899 GlySerGlyAlaGlyAlaThrThrAlaLySerSerGlyAlaProGluAspSerAspSerSer 918
Db 3168 GGGAGGCGGCGAGGCCACACAGCCCAAGAGCTGGGGCCCCCGGAGAGAGAGAGAGAG 3227
Qy 919 AlaThrCysSerAlaAspGluValAspGluAlaGluGluGlyGlyAspLyAspAlaGlyLeu 938
Db 3228 GCTAACCTGAGTGACAGCGAGTGATGAGAGCGGAGCGGAGCGGAGCGGAGCGGAG 3287
Qy 939 SerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAlaAlaSerProGln 958
Db 3288 TCCCAAGAGGCCAGGCTCTCAACCCGAGCTGGGAGCCCCGGGCGCAATGCTCAACCC 3347
Qy 959 LyAspProLeuAspLeuLyGluGluGluGluGluGluGluGluGluGluGluGluGlu 978
Db 3348 AAGCCACTGACCTGAAGAGCTGAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3404
Qy 979 ThrLyValAlaH1agluProProArgGluAspAlaAlaProThrLyAspProAlaProPro 998
Db 3405 ACCAAAGTCCATAGCCCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3464
Qy 998 aProProProProGluAsnLeuGlnProGluSerAspAlaProGluGlnProGlySerSe 1018
Db 3465 CCAACCGGCAAGCCCAAACTTGCAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3524
Qy 1018 rProArgGlyLySerSerArgSerProAlaProProAlaAspLyGlu----- 1033
Db 3525 CCCCCGGGCGAG 3584
Qy 1034 -----AlaPhaAlaAlaGluAlaGluGluGluGluGluGluGluGluGluGluGlu 1050
Db 3585 GTTCTTCCCAAGCTTGCAGGCCAGAGGCCAGAGAGCTGCTGGGAGACCCCTTGTGTGAG 3644
Qy 1050 rSerGlyLeuProPhaProValProProArgGluValIleLyAlaSerProH1saIaPr 1070
Db 3645 TTCCGGGCTGCTTCCCGGTGCCCCCGGTGAGGTGATCAAGAGCCCTCCCGGATGCC 3704
Qy 1070 oAaPProSerAlaPhaSerThrAlaProProGlyH1aProLeuProLeuGlyLeuH1sa 1090

Db 3705 GAGCCCTCAGCCTTCTCTACGCTCCAGCTGTACCCACCTGCCCCGGGCTCATGA 3764
Qy 1090 pThrAlaArgProValLeuProArgProProThrIleSerAspProProProLeuIleSe 1110
Db 3765 CACTGCCCGGCGCTCTGCGGCGCCACCAACATCTCCAAACGCGCTCCCTCATCTC 3824
Qy 1110 rSerAlaLyH1aProSerValLeuGluArgGlnIleGlyAlaIleSerGlnLyMeSe 1130
Db 3825 CTCTGCCAAGACACCCAGCGCTCTCGAGAGGCAAAATGATGCTCATCTCCAAAGAAATG 3884
Qy 1130 rValGlnLeuH1aValProTySerGluH1aAlaLyAlaProValGlyProValThrMe 1150
Db 3885 GGTTCACCTCCAGCTCCGTAATCAGAGATGCCCAAGGCCCGGTGGGCTCTGCACAT 3944
Qy 1150 rGlyLeuProLeuProMetAspProLyLyLyLeuAlaProPhaSerGlyAllyLySGln 1170
Db 3945 GGGGCTGCCCCCTGCCATGAGACCCCAAAAGCTGGACCCCTTCAGCGAGAGAGAGAG 4004
Qy 1170 uGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyAlaProThrAlaGlu 1190
Db 4005 GAGCTGTCCCAACGGGCGAGGCTGGCCACCGAGAGAGCTGGGGGTGCCACAGCCCA 4064
Qy 1190 nGluAlaSerValLeuArgGlyThraAlaLeuGlySerValProGlyGlySerIleThrLy 1210
Db 4065 GAGAGCGTCCGTCTGAG 4124
Qy 1210 rGlyIleProSerThrArgValProSerAspSerAlaIleThrThrArgLySerIleTh 1230
Db 4125 AGGCATTCACGACACAGGGGTGCCCTGACAGAGCCATCAACATCCGCGCTCATCAC 4184
Qy 1230 rH1SGlyThrProAlaAspValLeuThrLyGlyGlyThrIleThrArgIleIleGlyGluAs 1250
Db 4185 CCAAGGAGAGCCAGCTGACCTCTGTACAGAGGACCATACCAAGATCATCCGAGAGAG 4244
Qy 1250 pSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLyGlyH1aValIleTy 1270
Db 4245 CAGCCGAGTCCGTTGGACCGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4304
Qy 1270 rGluGlyLyLyLyGlyH1aValLeuSerTyArgGlyGlyLyMeSerValThrgInCySe 1290
Db 4305 CGAAGGCMAAGAGGCGACGCTTGTCTATGAGGGTGATGTGTGACCCAGTGTCTC 4364
Qy 1290 rLySGluAspGlyArgSerSerSerGlyProProH1SGluThrAlaAlaProLyAsrGln 1310
Db 4365 CAAG 4424
Qy 1310 rTyAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLe 1330
Db 4425 CTATGACATGATGAGAGGCGCGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4484
Qy 1330 uMetGlyArgAlaIleProProGluArgH1aSerProH1aH1aLeuGluGlnH1aH1 1350
Db 4485 CATGGCCCGGCTATCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4544
Qy 1350 sIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTy 1370
Db 4545 CATCCCGGGGTTCATACACAGAGATCCCTCGGTCTAGTGAAGGACAGAGAGACTA 4604
Qy 1370 rLeuArgArgGluAlaLyLyLeuLeuLyArgGluGlyThrProProProProProSe 1390
Db 4605 CTTGCGTCCGGAGGCGCAAGCTCTTAAGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 4664
Qy 1390 rArgAspLeuThrGluAlaTyThrGlnAlaLeuGlyProLeuLyLyLeuLyProAl 1410
Db 4665 ACAGGAGCTGACCGAGGCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4724
Qy 1410 aH1SGluGlyLeuValAlaThrValLySGluAlaGlyArgSerIleH1SGluIleProAr 1430
Db 4725 CATGAGGGGCTGTGTGACAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4784
Qy 1430 gGluGluLeuArgH1aThrProGluLeuProLeuAlaProArgProLeuLyGlyGlySe 1450
Db 4785 CGAGAGAGCTGGGACACAGCCGAGAGCTGCCCTTGGGCCCGGCGCTCAAGAGAGGCTTC 4844

1450 rIleThrGlnGlyThrProLeuIleuYrAaphrGlyAlaSerThrThrGlySerIleuYr 1470
Db CATCACGACGAGGACCCCGCTCAAGTACACACCGCGGCTCCACCACTGGCTCCAAAAA 4904
1470 gHiSaapValArgSerIleuIleGlySerProGlyArgThrPheProProValHisProLe 1490
Db GCACGACGTACCTCCCTCATCGGACGCCCGCCGACGCTTCCACCCGTGACCCGCT 4964
1490 uAapValMetAlaAapAlaArgAlaIleuGluAArgAlaCySyrGluGluSerIleuYrSa 1510
Db GGAATGTAATGGCGACGAGCGCGGACTGAAACGTGCTGTACGAGAGAGAGCTTGAAG 5024
1510 rArgProGlyThrAlaSerSerSerGlyIleSerIleAlaArgGlyAlaProValIleVa 1530
Db CGGGCCAGGAGCCGACGAGCTCGGGGGCTCCATTGGCGCGGCGCCCGCTTGT 5084
1530 lProGluLeuGlyIleProArgGlnSerProIleuThrTyrgluAapHisGlyAlaProh 1550
Db GCCTGAGCTGGGACAGCCCGGACAGCCCTTAACTATGAGACACAGGGGCAACCTT 5144
1550 eAlaGlyHisIleuProArgGlySerProValThiMetArgGluProThrProArgLeuG 1570
Db TGCAGGCACTTCCACGAGAGTTGGCCGTGACACGCGGAGCCACGCGCGCTTGA 5204
1570 nGluGlySerIleuSerSerSerIleuAlaSerGlnAapArgIleuThrSerThrPro 1590
Db GAGAGGCAACCTTGTCTCCAGCAAGGCATCCACGACCGAAAGCTGACCGCTGCA 5264
1590 gGluIleAlaIleYrSerProHisSerThrValProGluHisIleProHisProIleSerPr 1610
Db TAGAGTGGCGAAGTCCCCGACAGCACCGTGGCCGAGCACCCACCCCTTCGCGC 5324
1610 oTyrgGluHisIleuLeuAArgGlyValSerGlyValAapLeuTyrgIleSerHisIleProle 1630
Db CTATGAGCACTGTCTTGGGGCGTGAAGTGGAGCTGTATCGAGGCCACTTCCCT 5384
1630 uAlaPheAapProThrSerIleProArgGlyIleProLeuAapAlaAlaAlaIleTy 1650
Db GGCTTTCGAGCCGACTTCATACCCGCGGCATCCCTGTGAGCGAGCGGCTCTACGA 5444
1650 rLeuProArgHisIleuAlaProAapProThrTyrgProHisIleuTyrgProProTyrgIleu 1670
Db CCGCCCCACACACTGGCCCCCAACCCCACTTACCCGCACTGTACCCACCTTACCTCAT 5504
1670 eArgGlyTyrgProAapThrAlaAlaIleuGluAapAArgGlnThrIleIleAapAspTyrg 1690
Db CCGGGCTACCCCGACACGCGCGGCTGGAGAACCGGACAGACCATCATATGACTTACAT 5564
1690 eThrSerGlnGlnMetHisHisAapThrAlaThrAlaMetAlaGlnArgHisAapMetIle 1710
Db CACCTCGACGACATGCAACAAACGGGCAACGCGCATGCGCCACAGGAGCTATATGCT 5564
1710 uArgGlyLeuSerProArgGluSerSerIleuAlaLeuAspTyrgAlaAlaGlyProArgG 1730
Db GAGGGGCTCTCCCGCGGAGTCTCTGCTGGCACTCACTACGCTGCGGGTCCCGAGG 5624
1730 yIleIleAapLeuSerGlnValProHisIleuProValIleuValProProThrProGlyTh 1750
Db CATTCAATGACCTGTCCCAAGTGCACACCTGCTGTGCTGCCCCGACACAGGAC 5684
1750 rProAlaThrAlaMetAapArgLeuAlaTyrgLeuProThrAlaProGlnProPheSerSe 1770
Db CCAAGCACCGGCATGAGACGCGCTTGGCTACTCCCAACGCGGCCCTTACGACG 5744
1770 rArgHisSerSerSerProLeuSerProGlyIleProThrHisIleuThrIleYrProThrTh 1790
Db CCGCCACAGAGCTCCCACTCTCCACAGAGGTCCAAACACACTTGCACAAACCAAC 5864
1790 rThrArgSerSerGluArgGluArgAapArgAapArgGluArgAapArgAapArgGluArg 1810
Db CAGCTCTGCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5924

1810 gGluYrSerSerIleuThrSerThrThrValGluHisAlaProIleThrArgProG 1830
Db GAAAAAGTTCATCTCACGTCCACACGCGGTGAGACGACCACTGTGAGACTGG 5984
1830 yThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSer 1850
Db TACAGACGACGACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 6044
1850 gProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAapAl 1870
Db CCGCGCTCCCATCCCATGCCACAGCATCTGCCCATCTCCCTCGGACCGAGATGC 6104
1870 aLeuGlnGlnAapProSerValIleuHisAapThrGlyMetIleGlyIleIleThrAlaVa 1890
Db CCTCAGCAGAGACCCAGTGTCTTACAAACACAGCAATGAAGGAGGTATCATACCGCT 6164
1890 lGluProSerIleProThrValIleu-----ArgSerThrSerThrSerProVa 1907
Db GAGGCCACGACCGCCACGCTCTGAGAGTGGGCAAGTTCACCTTCACTTCCACCGCT 6224
1907 lArgProAlaAlaThrPheProProAlaThrHisCySProIleuGlyIleThrLeuAapG 1927
Db TCGCCGGCTGCACATTCACACTGACCACTGACCCACACTGGGGGACCCCTGATG 6284
1927 yValTyrgProThrIleuMetGluProValIleuLeuProGlyAlaProArgValAlaA 1947
Db GGTCTTACCTTACCTCATGAGACCCGTCTTGTGCTCCAGAGAGGCCCCCGAGGTGCGCG 6344
1947 gProGluArgProArgAlaAapThrGlyHisAlaPheLeuAlaIleProProAlaArgSe 1967
Db GCAAGAGCGGCCCCGAGCAGACCGGCGATCTTCTCTGCGCAACCCCGACCGGCTC 6404
1967 rGlyLeuGluProAlaSerSerProSerIleuGlySerGluProArgProLeuValProPr 1987
Db CGGGGCTGAGCGCGCTCTCCCGCAGCAAGGAGCTCGAGGCCCGCGCTGAGCTCC 6464
1987 oValSerGlyHisAlaThrIleAlaArgThrProAlaIleAapLeuAlaProHisHisAl 2007
Db TGTCTCTGGCCAGCGCACCATGCGCGCACCTCGTGAAGAACTTGCACCTTACCAACCC 6524
2007 aSerProAapProProAlaProProAlaSerAlaSerAapProHisArgGlyIleThrG 2027
Db CAGCCCGGAGCGCGCGCCCACTGCTCGGCTCGGACCGGACCGGAGGAGGAGCTCA 6584
2027 nSerIleProPheSerIleGlnGluIleuGluIleuAArgSerIleuGlyTyrgIleGlySerSe 2047
Db AGTAAACCTTTTCATCAGAACTGGAACCTCGCTCTGAGTTACACGCGCAGCAG 6644
2047 rTyrgSerProGluGlyValGluProValSerProValSerSerProSerIleuThrHisA 2067
Db CTACAGCCCCGAAAGGGGTGAACCCGTCAAGCCCTGAGACTCACCCAGTCTGACCCACA 6704
2067 pIleGlyLeuProIleuHisIleuGluIleuAapIleuAapIleuAapIleuAapIleuA 2087
Db CAAAGGCTTCCCAACACCTGGAAGAGCTCCAAAGAGCACTGGAAGGAGGAGCTGG 6764
2087 gProIleGlnProGlyProValIleuGlyIleGlyAlaAlaHisIleuProHisIleuA 2107
Db GCCAAGCAGCAGGCGCGCTGAGAGCTTGGCGGAGGCGCGCCACCTCCACACTGCG 6824
2107 gProLeuProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValI 2127
Db GCGGCTGCTGAGAGCAGACCTTGTGTCAGCCCGCTGTCAGACCGCCCGCAGAGGAGTCA 6884
2127 gGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAapTy 2147
Db AGGTACACAGCGGAGTGTACCTTGGCCAGCAATCATGAGTATACACAGAGACTA 6944
2147 rThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrgPhePr 2167
Db CACCGGACACACCCACAGAGTCAAGCAACCCGCGCGCGCCCTTACTCTTCC 7004
2167 oGlyAlaSerCySProValIleuAapLeuAArgArgProProSerAapLeuTyrgIleuProPr 2187

Db 7005 TGGGGCCAGCTGCCCCCTCCCTCCGACCTCCCGCCGACCAAGTACCTTACCTCCGCC 7064
Qy 2187 OPRoAAPHIaGlyAlaProAlaArgLysProHisSerGlyGlyLysArgSerPr 2207
Db 7065 CCGGACCAATGAGTGGCCCGGCGGTGCTCCCGACAGCAAGGGGCAAGAGTCTCC 7124
Qy 2207 OGluProAsnLysThrSerValLeuGlyGlyGlyLysArgGlyLysLeuProValSerPr 2227
Db 7125 AGAGCCAAACAAGACGTGCTTGGGTGGTGGTGAAGAGGATTTAACTGTGTCCC 7184
Qy 2227 OPRoGlyLysMetThrGluProGlyHisSerArgSerAlaValThrProLeuLeuThr 2247
Db 7185 ACCGGAGGCGATACGAGAGCCAGGCGACTCCCGAGTGTGTGAACCGCTGCTGACG 7244
Qy 2247 GAAPGlyGlyGluThrGluProSerArgMetGlySerLysSerProGlyLeuThrSerG 2267
Db 7245 GGATGGGGAAACAAGCAGAGCCAGAGATGGCTCCAAAGTCTCCAGCAACACACCA 7304
Qy 2267 nPRoProAlaPheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLys 2287
Db 7305 GCGGCCAGCTTCTTCCAGCAAGCTGACCGAGAGCACTCCGCCATGCTCAAGTCCAA 7364
Qy 2287 GGLngLysLeuAsnLysLeuLeuThrHisAsnArgAsnGluProGlyThrAsnLys 2307
Db 7365 GCAAGAGATCAACAAGAACGCTGAACCCCAACCGGAATGAGCTGAATACAAATAC 7424
Qy 2307 rGlnPRoGlyThrGluLeuPheAsnMetProAlaLeuThrGlyThrGlyLeuMetThr 2327
Db 7425 CCAAGCTGGAGCAGATGTTCAATATGCTCCCGCTACACGGAACAGGCTTATGACCT 7484
Qy 2327 rArgSerGlnAlaValGlnGlnLysLeuThrHisAsnArgAsnGluProGlyThrAsn 2347
Db 7485 TAAACCCAGCCGCTGCGAGAACATGCGACACCAACATGGGCTGAGGCCATTAAT 7544
Qy 2347 GlyValaLeuMetGlyLysThrArgGlnThrGluLysSerProProLeuSerAlaAsn 2367
Db 7545 AAAGCACTCATGGTAATATGACCAAGTGAAGAGTCCCGCGCTCAGCCGCAATGC 7604
Qy 2367 aPheAsnProLeuAsnAlaSerLysLeuProAlaAlaMetProLeuThrAlaAsn 2387
Db 7605 TTTTAACTCTGTAATGCGAGCTGCGCTGCTGCTATGCTCCCAATACCTGCTGA 7664
Qy 2387 rGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyValaLysValSerG 2407
Db 7665 CGGACCGAGTGAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7724
Qy 2407 yArgProSerSerArgLysValaLysSerProAlaProGlyLeuAlaSerGlyAspArg 2427
Db 7725 CAGAACCCAGAGCCGAAAGCCAAAGTCCCGGCGCGGCTGCAATCTGGGAGACGCG 7784
Qy 2427 OPRoSerValSerSerValHisSerGlyLysArgPheAsnArgThrProLeuThrAs 2447
Db 7785 ACCCTGTCTCTCTCAAGTGAACAACAACAACAACAACAACAACAACAACAACAACA 7844
Qy 2447 nArgValThrGluAspArgProSerSerAlaGlySerThrProPheProThrAsnPro 2467
Db 7845 CCGCGTGTGGAGGAACAAGCCCTGCTGCGAGGTTCCAGCCATTCCTCCACAAACCCCT 7904
Qy 2467 vLleMetArgLeuGlnAlaGlyValaMetAlaSerProProProProGlyLeuProAla 2487
Db 7905 GATCATGCGGCTGCAAGCGGCTGCTCAATGCTTCCCAACCCCGAGGCTCCCGCG 7964
Qy 2487 ySerGlyProLeuAlaGlyProHisHisAlaThrAspArgGluProLysProLeuLeu 2507
Db 7965 CAGCGGCGCTGCTGCGGCGCTGCGGCGCTGCGGCGCTGCGGCGCTGCGGCGCT 8024
Qy 2507 sSerGlnThrGluThrLeuSerAspSerGlu 2517
Db 8025 CTCGAGTACGAGACACTCTCCACAGCGAG 8055

RESULT 2
US-09-949-016-3804

/ Sequence 3804, Application US/09949016
/ Patent No. 681239
/ GENERAL INFORMATION:
/ APPLICANT: VENTNER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CLO01307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ. ID NOS: 207012
/ SOFTWARE: FASTSEQ for Windows Version 4.0
/ SEQ. ID NO 3804
/ LENGTH: 7912
/ TYPE: DNA
/ ORGANISM: Human
/ US-09-949-016-3804

Alignment Scores:
Pred. No.: 9,68e-183 Length: 7912
Score: 4165.50 Matches: 1103
Percent Similarity: 53.57% Conservative: 344
Best Local Similarity: 40.84% Mismatches: 777
Query Match: 31.52% Indels: 477
DB: Gaps: 109

US-09-522-753-5 (1-2517) X US-09-949-016-3804 (1-7912)

Qy 16 GluProArgThrProProHisSerLeuSerThrProValGlnLeuAlaGlyThrHisThr 35
Db 289 CAAGTCTGTTATCT 348
Qy 36 AspValGlyLeuLeuGlnLysGln-----HisHisSerArgAspThrAlaSerHisLeu 53
Db 349 GAGTTCGACAGCTCCCTATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 408
Qy 54 SerProGlySerLeuLeuGlnProGlnArgArgArgProSerLeuLeuSerGluPheGln 73
Db 409 TTCACACAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 468
Qy 74 ProGlyAsnGlnArgSerGlnGlnLeuHisLeuArgProGluSerHisSerThrLysPro 93
Db 469 CCAAGTTCGACAGCGCTCAAGAA-----AGGAGAACTAATTATGAAACCGTTTCATCCA 522
Qy 94 GluLeuGlyLysSerGluMetGluPheLeuGlnSerLysArgProArgLeuGlnLeu 113
Db 523 GGGCCATCCCAAGTATGATATTCATGATTCAGATTCAGATTCAGATTCAGATTCAGATTC 582
Qy 114 ProAsp-----ProLeuLeuArgProSerPro 122
Db 583 TCGATTCATTCATTCAGCGGTGTCAGTGCAGGTTTGGCTTTAATGACACCGGCTGCCA 642
Qy 123 LeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThr 142
Db 643 -----GAAGGCTGAGGCT--TCTGAGATGCTTAAGAAAGATCCAGCATTCGGA 690
Qy 143 GlyLysLeuGlu---ProValSerProProSerProProHisThrAsnProGluLeuGlu 161
Db 691 GCGAAACATGAGCTCACTCTCTCTCAATTCGGGCAACATGAGATGATGATCAAAAT 750
Qy 162 LeuValProProArgLysSerLysGlnGlnLeuLeuGlnAsnMetAspArgValaAspArg 181
Db 751 GCTTCACTTCAAAACCTCAAGGAAAGTTAATACAGATGATGATGATGATGATGATGAT 810
Qy 182 GluThrMetValGluGlnGlnLysSerLysLeuLysLysLysGlnGlnGlnGlnGln 201
Db 811 GAATTTGCAAAAGTAAACAGCAGATCTTAAATGAAAGAAAGAAACAAACAGGTTGAA 870

Db 2788 ---AGTAGAAGGTGGAACCTTAGATGAAGATTGGTGTAGCTCAGCAATAATGAATGCC 2844
Qy 912 -----ProGln---AepSerAepSerSerAlaThrCysSerAlaAepGluVal 926
Db 2845 CAAAGGCCCGGAGCCCGATCAGACAAATGATTTCCAGTGCACGCGCTGTATGAG--- 2901
Qy 927 AepGluAlaGluGlyGlyAapLyAaenAArgLeuLeuSerPro-----ArgProSer 943
Db 2902 ---GATGTGATGTGAGAGCAGAGAGAGAAATGTTTCTTATGAGACTCAAGACCTTCA 2958
Qy 944 LeuLeuThrProThrGlyAapProAlaAlaAenAlaSerProGlnLys-----ProLeu 961
Db 2959 CTGTTAAACCCCACTGGATCTTAATCTGTC---TCATCTCCGTTAAACCAATCCACTG 3015
Qy 962 AepLeuLysGlnLeuLysGlnAArgAlaAlaAlaLeuProProLys----- 976
Db 3016 GATCTGCCACAGCTTACAGATCGAGCTGTGTTATCCACCAATGGTATCTGCACCCCA 3075
Qy 977 -----GlnValThr 979
Db 3076 TGTAAACATACCAATTGGAACCCAGTAGCGGCTATGCTCTCTACAGGAGACATTAA 3135
Qy 980 LysValAlaGluProProArgGluAepAlaAlaProThrLysProAlaProProAlaPro 999
Db 3136 GCATGTGCATGAGTACAGCACTCTCTGAG----- 3162
Qy 1000 ProProProGlnAaenLeuGlnProGlnSerAepAlaProGlnGlnProGlnLysSerPro 1019
Db 3163 ---GACCACGGGAGAGACAAAGAACATGATTTGGAATGTAAGATTCTTCAAACTTCA 3219
Qy 1020 ArgGlyLysSerAArgSerProAlaProProAlaAapLysGluAlaAepAlaAlaGluAla 1039
Db 3220 TGTGGACATCCAGAGTCCA-----AACAGAGAG----- 3249
Qy 1040 GlnLysLeuProGlyAapProProCysTrpThrSerGlyLeuProProProAlaProPro 1059
Db 3250 -----TGS----- 3252
Qy 1060 ArgGluValAlaLysAlaSerProAlaAlaProAapProSerAlaAepSerTrpAlaPro 1079
Db 3253 ---GAAGTCTTCAAGCTCTCCACAT-----CAAGTGAATCAATCTTCCCT 3297
Qy 1080 ProGlyAlaAepProLeuGlnLysAlaAepThrAlaAArgProValLeuProArgPro 1099
Db 3298 GAAGGCGTTCCGGTTCCG-----ACAACTGACCAACAGGCGCCACCGCC 3342
Qy 1100 ProThrLysSerAapProProProLeuLysSerSerAlaLysAepProSerValLeuGln 1119
Db 3343 CCTCTCATCCCGTCAATCCAAACCAACAGTGGCTTCAGAAAA---CCATCTTTTATA--- 3396
Qy 1120 ArgGlnLysGlyAlaLysSerGlnGly-----MetSerValGlnLeuHis 1134
Db 3397 ---ATGGAGGGCTCCATCTCACAGGGAACACAGCACTTAATTGACTTCTCATTAATCAG 3453
Qy 1135 ValProLysSerGlyLysAlaLysAlaPro---ValGlyProValThrMetGlyLeuPro 1153
Db 3454 GCTTCTTACATCAGAAACACCCCAAGCCGTCACTGATCTATCTCTTGGACCTGCCA 3513
Qy 1154 LeuProMetAapProLysLysLeuAlaProPheSerGlyValLysGlnGlnLeuSer 1173
Db 3514 CGGCAACAGAAATCTGCCAAATCAGCTACTTGGCTTACATCAACAGCAAGAAATTTCT 3573
Qy 1174 ProArgGlyGlnAlaGlyProProGlnLysLeuGlyValProThrAlaGlnGluAlaSer 1193
Db 3574 CCCCGAAGCCAAACTCACACACTGAGGCTGTGTGTC---AGGCGCCAAATGAAAGT 3630
Qy 1194 ValLeuArgGlyThrAlaLeuGlySerValProGlyLysSerLysThrLysGlyLeuPro 1213
Db 3631 GTAGTCAAGAGTACCGCA---GGAGCCATACAAAGAAAGATTAATCTGGGGAATCTCA 3687
Qy 1214 SerThrArgValProSerAapSerAlaAlaThrTrpArgGlySerLysThrLysGlyThr 1233
Db 3688 ACCAGCAAAATTTCAGTGAGAGCATTCATCTCCCTACGGGGGCTCTATCACTCAGGGCACCC 3747

Qy 1234 ProAla-----AepValLeuTrpLysGlyThrLysThrArg 1245
Db 3748 CCGGCTCTTCCCGCCAGCTGGCATACCAAGAGGCTTGTGTGAAGGGTCCATTTGAGA 3807
Qy 1246 IleIleGlyLysAapSerProSerAlaAepAlaGlyAArgGluAapSerLeuProLys 1265
Db 3808 ATGCCCATTTGAAGACAGCAGCTCT-----GAGAAAGGACAGAGAGCACTGCATCCAA 3861
Qy 1266 GlyHisValIleThrGlyGluGlyLysLysGlyHisValLeuSerTrpGlnGlyMetSer 1285
Db 3862 GGCCATGTTATTTAAGAAAGCAAAAGTGAACATATCTTCAATATATATTAAGAAT 3921
Qy 1286 ValThrGlnCysSerLysGluAepGlyAArgSerSerGlyProProIleGluThrAla 1305
Db 3922 GCC-----CGAAGAGGACTAGAGTCCAGAAACAGCTCATGAATCAAGT 3966
Qy 1306 AlaProLysArgThrTrpAepMetMetGluGlyAArgValGlyAArgAlaLysSer 1323
Db 3967 TTA---AAGAGAAGCTATGATCAGTGAAGAAATATTAAGCAAGGATGTCAATGAG 4023
Qy 1324 -----SerAlaSerLysGlnGlyLeuMetGlyAArgAlaLeuProProGluArg 1339
Db 4024 GAGTCTCTGTATCAGCACCGTTAGAGGGCTGATATGCGGACATTAACC-----AG 4077
Qy 1340 HisSerProHis---HisLeuLysGlnGlnHisIleAArgGlySerLysThrGlnGly 1358
Db 4078 GGAAGTCTCTATTCGACTCCAAAGAAAGACTGATTTCTCGGCTCCATATGCAAGGG 4137
Qy 1359 IleProArgSerTrpValGluAlaGlnLysAepTrpLeuAArgGluAlaLysLeuLeu 1378
Db 4138 ACACCAAGAGCAACAACATAAGCTTTGAAGATGCTT---AAATATCCCAACAATTT 4194
Qy 1379 LysArgGluGlyThrProProProProProProSerAArgAepLeuThrGluAlaTrpLys 1398
Db 4195 AAAAGGGAAGT-----CCTCCATACGAGCATTTGAAAGTGCATTT--- 4236
Qy 1399 ThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGlnGlyLeuAlaThrVal 1418
Db 4237 -----ACCAAGGAACCA---TATGATGCG---ATCACACCATTC 4272
Qy 1419 LysGluAlaGlyAArgSerLysIleGlnLysProArgGlnGlnLeu----- 1433
Db 4273 AAAGAAATGGGCGTTCCATTCATGATTCGAAGCAAGATATTTAATCAGAAAGT 4332
Qy 1434 ArgHisThrProGlnLeuProLeuAlaProAArgProLysGlnGlySerLysThrGln 1453
Db 4333 CGAAAACTCCAGAAAGTGTCCAGAGCACACGCGCATTAATGAGGTTCCATTCCAG 4392
Qy 1454 GlyThrProLeuLysTrpAepThrGlyAlaSerThrThrGlySerLysLysHisAepVal 1473
Db 4393 GGCACACCAATTAAGTTTGACAACAAC---TAGAGTCAATTCGCCATTCACCAATGTC 4449
Qy 1474 ArgSerLeuIleGlySerProGlyAArgThrPheProProValHisProLeuAapValMet 1493
Db 4450 AATCTCTTAATACCGGGGCTTGCAAACTATCCCGGAAGCCCTCCGTGAAATATGG 4509
Qy 1494 AlaAap---AlaArgAlaLeuGlnLysGlyAcyTrpGlu-----GlnSer 1507
Db 4510 CCAGGAACATTAAGGTGTAGAACGGGGAATATGAGATGTGAAGACGCGAGACC 4569
Qy 1508 LeuLysSerArgProGlyThrLysSerSerSerGlySerLysAlaArgGlyAlaPro 1527
Db 4570 GTGCTTCCCGGACACGTCAGTGAAGCTTGGCCCTTCCTTTAGGTCCACA--- 4626
Qy 1528 ValIleValProGlnLeuGlyLysAepProArgLysSerProLeuThrTrpGlnAapHisGly 1547
Db 4627 -----CTGCATGAAGCTCCCAAGACAACATGAGCCCTGGGATTTATGATACACAGCT 4680
Qy 1548 Ala-----ProPheAlaGlyHisLeuProArgGlySerProValThrMet 1562
Db 4681 GCACGAGAGACCCCTGTGATGATTAACAACACATGCTCAGAGGCTCACCCATG----- 4734

Oy 1563 ArgGluProThrProArgLeuGlnGlySerLeuSerSerLyala---SerGln 1581
 Db 4735 -----ATGAACAGAACTTCGATGTACAAATTTCTTCAACAGACTCCAAATCAT 4785
 Oy 1562 AspArgLyLeuThrSerThrProArgLys-----LleAlaLysSerProHis 1597
 Db 4786 GAAGAGAAATGACACGTACGCCCTTACCCAGAGGAAAGTATCCAGCGAAAGCTCCAGTG 4845
 Oy 1598 SerThrValProGlnHisIleProHisProLysSerProTyGlnHisIleLeuArgGly 1617
 Db 4846 CCGGGGGTGAACCCCTGCTGAGCCAC-----ACTCCCTTGAATCCCATACAGAGGC 4899
 Oy 1618 ValSerGlyValAspLeuTyArgSerHisIleProLeuAlaIleAspProThrSerIle 1637
 Db 4900 AGCACTGACAGGAGTTTATCGAGCCACCTGCCAGACACTTGATCA-----GCCATG 4956
 Oy 1638 ProArgGlyIleProLeuAsp---AlaAlaAlaAlaTyTyTyLeuProArgHisIleAla 1656
 Db 4957 CTTTTTCACAGGGCTTTGATCTCTGACGGCTGCTTCACTGTTTCAAGAGACAGCTTCA 5016
 Oy 1657 ProAspProThrTyProHisIleuTyTyProTyTyLeuLleArgGlyTyProAspThr 1676
 Db 5017 CCAACTCCAGGTACCCCAAGTCAGATCAGCTTAC----- 5052
 Oy 1677 AlaAlaLeuGluAsn---ArgGlnThrIleIleAsnAspTyTyIleThrSerGlnIleMet 1695
 Db 5053 ---GCAATGAGAAACACAGACAGACAAATCTTAATGATTAATCACTCAACACAGATG 5109
 Oy 1696 HisHisAsnThrAlaThrIleMetIleGlnArgAlaAspMetLeuArgGlyLeuSerPro 1715
 Db 5110 CAAGTAACTTG-----CGTCCAGATGTGGCCAGAGAGACTCTCCCA 5151
 Oy 1716 ArgGluSerSerLeuAlaLeuAsnTyTyAlaAlaGlyProArgGlyIleIleAspLeuSer 1735
 Db 5152 AGAGAGCAGCCACTGGGTCTCCCATACCCAGCA---ACGAGAGATATCATTAACCTGACC 5208
 Oy 1736 GlnValProHisIleuProValIleValProProThrProGlyTyThrProAlaThrIleMet 1755
 Db 5209 AATATGCTCCCA---ACAATTTTATGCTTCATCCAGGGGGAACAGCACTCTCCCATG 5265
 Oy 1756 AspArgLeuAlaTyTyLeuProThrAlaProGlnProHisSerSerArg---HisSerSer 1774
 Db 5266 GACAGAAATCACTTAATTCCTGCTGACACAGATTACTTCTCCAGAGCGGTACAACTCT 5325
 Oy 1775 SerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSer 1794
 Db 5326 GCTTCCAGTCTCCCAAGACCCCAACACACTT-----GCACTGCTGCAAGTCT 5376
 Oy 1795 GluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGlyLysSerIle 1814
 Db 5377 GAGAGGGAAACGGGAACGGGAGCGGAGGAAGCGGAGCGGGAACGG-----ATT 5427
 Oy 1815 LeuThrSerThrThrThrValGlnHisIleAlaProIleTPArgProGlyTyThrGlnIleSer 1834
 Db 5428 GCTGCAGGTTCTCTCGACCTC-----TACTGCGCGCCAGGCTCAAAAG--- 5472
 Oy 1835 SerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgProLaserHis 1854
 Db 5473 -----CTGGCGGACCTGGCAGTCA 5493
 Oy 1855 SerHisAlaHisGlnHisSerProLysSerProArgTyGlnAspAla---LeuGlnGln 1873
 Db 5494 GGATATGTTCCGCTCCCTTCCCT---TCAGTAAGAACTCAGAGACCATGTGTCAAACG 5550
 Oy 1874 ArgProSerValIleuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSer 1893
 Db 5551 AGACCCAGTGTGTTTCCAAAGAAACCAATGCAACCACTGTATATCACCTTGGATCAACT 5610
 Oy 1894 LysProThrValIleuArgSerThrSerThrSerSerProVal-----ArgPro 1909
 Db 5611 GCTACGCTACGAATATGCACTGCTGCTGGGGGCGCTTCAATAAGCAAGCGCTGCA 5670
 Oy 1910 AlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyTyThrLeuAspGlyValTy 1929

Db 5671 GCTTCCCGCTTCAACACTGCTGCGGATGCC---CTGGCTGCTTGTGGATGCTGCACT 5727
 Oy 1920 ProThrIleuMetGluProValIleuLeuProLys-----GluAlaProArgVal 1945
 Db 5728 TCTGCACCCCAAGATGATGTGTCCAAAACAAAGAGATTAAGATGAAGCTGCCAGGTTA 5787
 Oy 1946 AlaArgProGluArgProArgAlaAsp----- 1954
 Db 5788 GAAGAAATTTGAGAAAGAGGTGACAGACAGTGTAGAACAGACAGACTAGACAGAA 5847
 Oy 1955 -----ThrGlnHisAlaPheLeuAla 1961
 Db 5848 ACCCTGAGGTGAGAAAGATCTGTTCAGTGTATTATACCTTTCTTCAAGCTTTCAGT 5907
 Oy 1962 LysProProAlaArgSerGlyLeuGluProLaserSer-----ProSerLyGly 1978
 Db 5908 GGCAAGCCC-----CAGCTCATTTCTTCAAGTATTATTCGAGGCTGG 5952
 Oy 1979 SerGluProArgProLeuValProProValSerGlnHisAlaThrIleAlaArgThrPro 1998
 Db 5953 AAAGTAAAGGGCT-----CTCCAAATTCAGATATGAGAAAGCTAAAGCACAGA 6006
 Oy 1999 AlaLysAsn-----LeuAlaProHisIleAlaSerProAspProProAlaProProAla 2016
 Db 6007 GAGAAAGTACCATTAATTCAGCTAATTCATAGACGTATATCACCAGGCAAAATTGCC 6066
 Oy 2017 SerAlaSerAspProHisIleArgGlyLeuThrGlnSerLysProHisSerIleGlnGluLeu 2036
 Db 6067 TCGACAAAGATTCGAGGAAACGTGGCTCMAAGTTACAGCTCTTC----- 6114
 Oy 2037 GluLeuArgSerLeuGlyTyHisIleGlySerSerTyTySerProGluGlyValGluProVal 2056
 Db 6115 -----AGTACATTAATCTTTCACAGATGATAACACCTTACAGATGATGAGGTGATA 2076
 Oy 2057 SerProValSerSerProSerLeuThrHisIleAspLysGlyLeuProLysHisIleGlnGlu 2076
 Db 6169 AGCTCTGACACTCACTGCGGCCCAACCCAGAGAAATGCGACAGCTTATCAACCAAGGTT 6228
 Oy 2077 LeuAspLysSerHisIleuGlnGlyLeuArgProLysGlnProGlyProValLysLeu 2096
 Db 6229 GTTAAGGCAATTCAGCGGAAATGATCTTACACAGAAATGAAAGAC--- 6279
 Oy 2097 GlyGlyGluAlaAlaHisIleuProHisIleuArgProLeuProGluSerGlnProSerSer 2116
 Db 6280 -----TTACATCATATGACCA-----CAGCAGAAATCACCA 6312
 Oy 2117 SerProLeuLeuGlnThrAlaProGly-----ValLysGly 2128
 Db 6313 TCTCCCAACAAACAGCTGCCCTTCTTACAGAGCAGAGGAAATGGGCAAGTCCCAAG 6372
 Oy 2129 HisGlnArgValIleThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyThr 2148
 Db 6373 ACCCATGCGGTGATCACTGTTGATATCAATCTGTCAAAATTAACCAAGATTTTGGCT 6432
 Oy 2149 ArgHisHis-----ProGlnLysSerAlaProLeuProAlaProLeu 2163
 Db 6433 AGAATCAAGTTTCTCGCAGATCCCAAGCAG-----CTCTTAATCTTACATTC 6483
 Oy 2164 TySerPheProGlyAla-----SerCysProValIleuAspLeuArgTyArgProProSer 2181
 Db 6484 CAGAACTCACCTTCTGCTTGTGATATCAACCTG-----AGGACTAAACATCA 6534
 Oy 2182 AspLeuTyTyLeuProProProAsp-----HisGlyAlaPro---AlaArg 2195
 Db 6535 AACCTTACAGCCCAAGATCCCAAGCTAGTGTGCACTCATCAAAAGACAGCTTCAAG 6594
 Oy 2196 GlySerProHis-----SerGlnGlyGlyLysArgSerProGluProAsn 2210
 Db 6595 GTCTCTCAGAAATCTTGTGAGCAAAATCAGGGGAAGTAGCGCTGAGAAATCCCAAG 6654
 Oy 2211 LysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGly 2230

Db 6655 AGAGTCACGCTC-----TCTTCGAGCCCTACGAGCCCATCTCCCAACCCAG--- 6702
Qy 2231 MetThrGluProGluHisSerArgSerAlaValTyProLeuLeuTyArgAspGlyGlu 2250
Db 6703 ---GTTCCGGTTGGCATGAGAAACAGACAGCTTCTGCTCTGTCTCAGAGGGGCGCA 6759
Qy 2251 GluThrGluProSerArgMetGlySerIleSerProGlyLeuLeuSerGlnProProAla 2270
Db 6760 GAGCTCTCAGACGAGAGAAATGATGCCGCTCACAGGAGATATAGTACTTCTCCTTCA 6819
Qy 2271 PhePheSerIleLeuThrGluSerAsnSerAlaMetValIleSerIleGlyGlnGlu 2290
Db 6820 TCTTTCACCAAGCTT---GAAATATACATCACCCATGGTTAATCAAGAGCCAGAGATT 6876
Qy 2291 AsnIleValSerLeuLeuThrHisAsnArgAsnGluProGluTyraenIleSerGlnProGly 2310
Db 6877 TTTTCGTAAGTTGAATCTCTCTGCGAGCGTACCTCTGATATGAGAGCTCTCAGCCAGGA 6936
Qy 2311 ThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyraSerGln 2330
Db 6937 ACTGAGATCTTAACTGACGAGCACTTACTACGTCAGGCTCAGTACTAGTACGAGGCGCAT 6996
Qy 2331 AlaValGlnIleHisAlaSerThrAsnMetGlyLeuGlnAlaIleIleGlyValAlaLeu 2350
Db 6997 TCTTTTGTCTATCCTGCCAGT---AATCTTGCGCTGAGACATTAATCAGAGAGCTCTC 7053
Qy 2351 MetGlyValTyraSerGlnTrpGluGlu-----SerProProLeuSer 2364
Db 7054 ATGGGAAGCTTTGATGCAAAAGTTGAGATCATGGAATGTGATGCTCCAGCCTATAGGA 7113
Qy 2365 AlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThr 2384
Db 7114 GTA-----GTGCTGTGACTGCCAACACCTCACTT-----GTACCC 7149
Qy 2385 AlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGly-----Lys 2402
Db 7150 AGTGGTGACACAGAGAGAGAGAGGAGCCCATCACTTATTCAGAGAGATTGCAAA 7209
Qy 2403 AlaIleValSerGlyArgProSerSerArgIleAlaIleSerProAlaPro-----Gly 2420
Db 7210 CCAAAGCTGATCGCAAGTCAAAACAGAGAAATCTAAAGTCTCTTATCTAGCTGGCAAGGC 7269
Qy 2421 LeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluGlyAsnPro 2440
Db 7270 TACTTAGAAGAGGAAAGCGCCCTCTTCACTCTCTCTATCATTCAGAGAGGATTAACAT 7229
Qy 2441 ArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThr 2460
Db 7330 AGGCAACGCGCA-----GGGTGGGCTCTGGGAAAGACAGGCCCTTTCACAGGCTCAACT 7383
Qy 2461 ProPheProTyraAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProPro 2480
Db 7384 CAGTTTCCTTATACCTCTGACTATGCGGATG-----CTCAGCACTACTCCACCA 7434
Qy 2481 ProProGlyLeuProAlaGlySerGly---ProLeuAlaGlyProHis-----His 2496
Db 7435 ACAACGATTCGATGATGCTCTCTGCGGTGACCAACAGCACTCTCTCAACAGAGACAGG 7494
Qy 2497 AlaTrpAspGluProGlyProGlyProLeuLeuCySerGlnTrpGluThrLeuSerAspSer 2516
Db 7495 ATCTGGAGGAGAGAGCTGCGCCCATGCTCTCAGACAGATAGAGACCTGTGCGATAGT 7554
Qy 2517 Glu 2517
Db 7555 GAT 7557

RESULT 3
US-09-144-085-3/c
; Sequence 3, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
; APPLICANT: Gueatafeson, Clae
; APPLICANT: Belach, Mary C.

APPLICANT: Ashley, Gary
APPLICANT: Julien, Bryan
APPLICANT: Ziemann, Rainer
TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
TITLE OF INVENTION: THEREFOR
FILE REFERENCE: 30062-20020.20
CURRENT APPLICATION NUMBER: US/09/144,085
CURRENT FILING DATE: 1998-08-31
EARLIER APPLICATION NUMBER: 09/010,809
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 33529
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-09-144-085-3

Alignment Scores:
Pred. No.: 9,66e-24 Length: 33529
Score: 726.50 Matches: 674
Percent Similarity: 32.36% Conservative: 314
Best Local Similarity: 22.08% Mismatches: 1199
Query Match: 5.50% Indels: 873
Gaps: 137

US-09-522-753-5 (1-2517) x US-09-144-085-3 (1-33529)

Qy 17 ProArgTyraProPro---HisSerLeuSerTyProValGlnIleAlaArgThrHisThr 35
Db 22686 CCAGGCGCTCCACGACGACCATGCTCCCGGCTGCGACCTTGCAGCA-----TCAGCC 22633
Qy 36 AspValGlyLeuLeuGluTyrglnHisSerArgAspTyraIleSerHisLeuSerPro 55
Db 22632 GACCTCGAG-----CGCACTAGTCTGCTGCTGCCGCA 22597
Qy 56 GlySerIleIleGlnProGlnArgArgProSerLeuLeuSerGlnPheGlnProGly 75
Db 22596 GGGTCAGACACCGCGCAATGGCAGCGCTCACTCTTATGAGT-----GCCCGAGCA 22540
Qy 76 AsnGluArgSerGln-----GluLeuHis----- 83
Db 22539 GCCACAGGGGCTCAGCCGCCACGATCCAGCGTACAGCGGCCACTTCCAGCGCAAA 22480
Qy 84 -----LeuArgPro---GluSerHisSerTyraLeuProGluLeuGly 96
Db 22479 CAGCCGGGCTCGGTGATCTCGTCCGCTCCAGCAAGCTCTTCTGCTGCCGCGGC 22420
Qy 97 LysSer-----GluMetGluPheIleGluSerTyraArgProArgLeuGluLeu 113
Db 22419 CGGGAACACACCTCTCTCAACCCAGGTGAGATAGCGCTCCAGCGCCTCGCACACTC 22360
Qy 114 ProAsp-----ProLeuLeuArgProSerPro 122
Db 22359 GTCGAAGCGCGCACGGAACAGCGGGTACACTTGTAAGCTCTTCCCTCCATCCGAGCCG 22300
Qy 123 LeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrTyraAspArgSerLeuThr 142
Db 22299 CTGGCTGCCCTG-----CCCTGTAACAGCAC 22273
Qy 143 GlyLeuLeuGlnProValSerProProSerProHisThrAspProGluLeuGluLeu 162
Db 22272 CGCAAGCTT---CCGCGCTGCTTCCGCGGCTCCGCTCAC-----CACCGCGCATC 22225
Qy 163 ValProProArgLeuSerIleGluLeuIleGlnAsnMetAspArgValAspArgGlu 182
Db 22224 CGGCGCGCGCGAGCAGCAGCGCGGAGACCTCTCCACAGCTCCCGCAGCGCTCCCGCAG 22165
Qy 183 IleThrMetValGlnGlnIleSerIleValLeuValSerGlnGlnIleGluGlu 202
Db 22164 CACCGAGCGCGCGGACTCGAATGCGCTCCGATGACAGCGCGCGCT-----GCG 22117
Qy 203 GluAlaIleValProProGluProGlu-----LysProValSerProProIle 219

Db 22116 CACCAAGTCCGACACCCCAACCTCCCGGCTCTTCCAGCACTTCCGACCCGCGCCG 22057
 Qy 220 GtuserlyshiaargserleuValglnleiletryaspgluasn-----Arglyshlys 237
 Db 22056 CTGCCATTCACCGCGGCTCTTCGACCCGACGACGACGAGCAATAGCGCAGCGCC 21997
 Qy 228 AlaGluaIaAlaIshArgIleLeuGlulysLeuGly----- 249
 Db 21996 CGCCTCGGCTCGACAGGCTCCGCGCGGCTCCGCGCGCTTCTTCAGAGATGATATG 21937
 Qy 249 ----- 249
 Db 21936 CGCGTTTCGTCGCGGTATCCGACGACGACGCGGACGACGAGCCGCGCGTTGCG 21877
 Qy 250 ---ProGlnValgluleuProleuTryaanglnProserAspThrArglnTryhlsGlu 268
 Db 21876 CGCCACGACGCGGCTCTTCGACGACGACGACGCGCT-----TCTTCCACCC 21826
 Qy 269 AsnIleLysIleAsn-----GlnAlaMetArgLysLeuIleuTryphe 284
 Db 21825 AATGTCGGGCTCGCTCGCGCTCGCGACGCGCTTCCGACGACCTCGCTGATGCA 21766
 Qy 285 LysArgArg---AsnIshIaArgLysGlnTryLysGlnLysPheCysGlnArgTryaap 303
 Db 21765 CAGACACATCTTGATCAGCGCCG----- 21742
 Qy 304 GlnLeuMetGlnAlaLeuGlnLysLysValGluArgIleGluAsnAsnProArgArg 323
 Db 21741 -----CACGCGCGCGCGCG 21727
 Qy 324 AlaLysGlnSerLysValArgGluTryTryGlnLysGlnPheProGlnIleArgLys--- 342
 Db 21726 CTGCGATGTCGAGGTTGCA-----CTTCACCATCC 21694
 Qy 343 -----GlnArgGlnLeuGlnGlnLysMetGlnSerArgValglYlnArgLysSerGly 360
 Db 21693 CAGGTACAGCGGACGCTCGGGGCTACGCTCGGTCMAACCTCGGACGCGCTCGGAG 21634
 Qy 361 LeuSer---MetSerIshIaArgSerGlnIshGlnValSerGlnIleIleAspGlyLeu 379
 Db 21633 CTGCGATCGGCTCTCAAGGCTCGTCCGCTACCGTACCGTCCGCGCTCGATGTCCTC 21574
 Qy 380 SerGlnGlnLysAsnLeuGlnLysGlnMetArgGlnLeuAlaValIleProPomLeu 399
 Db 21573 GGGCGACAGACCA-----CGACGACGCGCTCGCGGATCACCGCTGCTG 21526
 Qy 400 TyrAspAlaAspGlnGlnArg-IleLysPheIleAsnMetAsnGlyLeuMetAlaAspPr 419
 Db 21525 GGCAGGGCGCTTCGCGCGCTGACCGCTGCGACCGCTGCTTCACCGCAGAGCC 21466
 Qy 419 ometyLysValTryLysAspArglnValMetAsnMetTrpSerGlnGlnLysGln 439
 Db 21465 ACCGATCACCCCGACGACGCGTCCGCTCGCGCTCGCGCTGAGA-CAGCGCTTCAA-- 21409
 Qy 439 rPheArgGlnLysPheMetGlnIshProLysAsnPhagly---LeuIleAlaSerPhe 458
 Db 21408 -----CAACAGCATCCCGACACCTTCGCGGAGCATCCCTT-----GAGAC 21365
 Qy 458 uGlnArgLysTryhValAlaGlnCysValLeuTryTryLeuThrLysLysAsnGluAs 478
 Db 21364 GCACGAGAAAGCTCTTCGACCGACCGTCCGCGGAGCATCCCTT-----GAGAC 21317
 Qy 478 nTryLysSerLeuValArgArg-----SerTryArgArgArgGlyLysSerGlnGln 496
 Db 21316 GGCTGAATTCACGAAGCGCGGGGTGCTCATCACCTTCACCGCGCGCGCGGCCCA 21257
 Qy 496 n-----GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 509
 Db 21256 GGTTCGATTCGCGCTCGCGCGCGCGCTGACCGCGAGTGCAGGACGACGAGCAGCAG 21197
 Qy 510 -----GlnProMetProArgSerSerGlnGln 518

Db 21196 AGCAGCCGCTGTCACCGCTCATCGCTGAGCCCTGACGCCCA--CGCGTAGGCCACCC 21140
 Qy 518 uGlnLysAspGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGln 538
 Db 21139 GCGCCGAGAGATGCTCCCGCGCTCCCGATCCCGGATCCCGGATCCCGGATCCCGAT 21080
 Qy 538 IglLysAspLysGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 556
 Db 21079 CGACAGTATGAAGACACCGCTGCTCGGAGCCCATGAAACCGAGTAACTCCGTCGCGC 21020
 Qy 556 uAspAsnAspGlnLysGlnLysValAlaLysLys----- 567
 Db 21019 TCCCGCTCAGCGCCGACGCGCGACGCGCGGACCGCTCCCAATGCTCCACGCCCTTCCA 20960
 Qy 568 -----GlyArgLysTryhAlaAsnSerGln 575
 Db 20959 GCACACGCGATGCTCGCGATTCATGACTGCGCTCGCGGCGGAGTA-TCCGAAAGAG 20901
 Qy 575 nGlyArgArgLysGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 595
 Db 20900 TTGCGGTGAAACAGGTCGATGCCGAAACCAACCAACCTTCGCGCGCTTCGCT 20841
 Qy 595 aLethrProGlnGlnSer-----AlaGlnLeuAlaSerMetGln 608
 Db 20840 GCGGCTTCGATCAGGCTCGTACCTGAAACCGCTCCGACGCGCTCGGAGCGCTTCG 20781
 Qy 608 uLeuAsnGlnSerSerArgTryPheGlnGlnLysGlnLysGlnLysGlnLysGlnLysGln 628
 Db 20780 ATCGGCTCTTCCTCCCTCGCGCAAGAGCGCATGTCGCGGCTGCAACGCTTCC 20721
 Qy 628 uGlnIshGly-ArgAsnTrpSerAlaIleAlaArgMetValGlySerTryhValSerG 648
 Db 20720 GCGACCGCGACGATCATGACGATCGGATCGGCTCTCTTCCCTGACGACGCGC 20661
 Qy 648 IncLysLysAsnPheryrPheAsnTryLysLysArgGlnAsnLeu----- 662
 Db 20660 ACCCGCGCGCGCTTCTTCACTGACCGCGCGGAAAGCGCTTCTTCGACGACGCTCT 20601
 Qy 663 -----AspGlnIleGlnGlnIshLysLysLysLysLysLysLysLysLysLysLys 673
 Db 20600 GCGATCGCGCGCGCGCTCGGATGTCMAAGACGAGGTCGCGGAAAGACGTCGCTG 20541
 Qy 673 eGlnLysGlnLysAsnAlaArgArgLysLysLysLysLysLysLysLysLysLysLys 685
 Db 20540 CCGAGCTGAGCGGTCGCGAGCGCGACCGCATGACGAGTCGAGCCGACGCTCTTG 20481
 Qy 686 -----AlaProAlaAlaLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 698
 Db 20480 GCACCTGATCCGCTGACAGCTTCGCGCGCTGAAAGCTTCGACGCGCGACCTTCCG 20421
 Qy 698 roValValGlnAspGlnGlnGlnGlnLysSerGlyValSerGlyAsnGlnGlnGlnGln 718
 Db 20420 CCGGACCGAGTCGACGAGGATTCAGCGCGCTCCGCGCGCGCGCGAGCGCT 20361
 Qy 718 aGlnGlnLysGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 738
 Db 20360 CCGGAGCGCGCGAGGTCGCTGCTGCTGTCGCGAGGACCGCTTCGCGCAAG--CTCG 20307
 Qy 738 LysProAlaThrValAsnAsnSerSerAspThrGlnSerIleProSer-ProIshTrhGln 757
 Db 20306 GCGGACGAGCGACGAAACAGGCGCGAGCTCGCGCTGAGCTCAGCCACGCTGCA 20247
 Qy 758 AlaAlaLysAspThrGly----- 763
 Db 20246 GCTGCGGAGATTCAGTGCATCGGACCAAGCTGCTTCAGGCGTGAAGCGCGCGCT 20187
 Qy 764 -----GlnAsnGlyProLysProProAlaThrLeu----- 773
 Db 20186 CCAGCAGAGCGAGCGCTTTCACCGGATCGGACGAAACCATTCGCGCTCATGCGCG 20127
 Qy 773 ----- 773
 Db 20126 ACAGCTCGGCTTCGCCAAGTGCCTGATCCCAAGCGCGCGCTGTCAGAAAGCCC 20067

[illegible]

QY	1082	-----HisProLeuProLeuGlyLeuHisAspThrAlaAspProValLeuProArg	1098
DB	19061	GATGGCAATGCGCTCGGCTCTCAAGGCTT-----CGACCAAGCTGCTGCTC	1901
QY	1099	ProProThrIleSerAsnProProProLeuIleSerSerAlaValHisProSerValLeu	1118
DB	19013	CTTCGGCTGCACCGACCAACACCA-----::::-----CGAACCGGCTCCAGGGGGGCTG	1896
QY	1119	GLuArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSer	1138
DB	18965	CCGGGA-----GCTCCACAGAGCTGAAGCTCACCTGATATGATTGGCCACCAG	1891
QY	1139	GLuHisAlaValAlaProValGlyProValThrMetGlyLeu-----	1152
DB	18914	CGTGAACGGCGGTCCGACAGCTCTCGGCGCGTCCGCGACGATGTAGAGCCGATGC	1885
QY	1153	---ProLeuProMetAspProLysLysLeuAlaProPheSer-----	1165
DB	18954	TCGCCACAGGCTT-----GCCCGGTGAAGTCTCGACCTTGACCGAG	1881
QY	1166	-----GlyValLys-----::::-----GlnGlnGlnLeuSerPro	1174
DB	18812	CGAGTGTCTGGTGTGGCGCTCGCTCGTAGAGCTTACCCGGACCCGAAGCTCGCTCGCC	1875
QY	1175	ArgGlyGlnAlaGlyProProGlnSerLeuGlyAlaProThrAlaGlnGlnAlaSerVal	1194
DB	18752	CCGTGCGGTGACGGCCACGT-----CCGACACAGCAAAAGCGCACGAGA	1870
QY	1195	LeuArgGlyThrAlaLeuGlySerValProGlnGlySerIleThrLysGlyIleProSer	1214
DB	18707	CCTCGTTCG-----CCCTGGAGCTCTGGAGAAAGCGCGACCATCG	1866
QY	1215	-----ThrArgValProSerAspSerAlaIleThr---TyrArgLysSerIleThrHis	1231
DB	18665	TGTGGAGCGCGCGCTCATACAC-GCGGGGTGCACGGCATACGCTCGCGCTGCTCTC	1860
QY	1232	GlyThrProAlaAspValLeuTyrLysGlyThrIleThrArg-----Ile-I	1247
DB	18606	TCGGTTCCTCC-----GGACAGCACCCGACCGAAGTAGTGTGCTCT	1856
QY	1247	egLGLuAspSerProSerArgLeuAspArg---GLuArgGluAspSerLeuProLysG	1266
DB	18564	TGGGCGCAAGCTCTCAAGAGCCCTTGAAAGCGCGGGCGGT-----	1852
QY	1266	YHisValIleTyrGlnGlyLysGlyHisValLeuSerTyrGlnGlyGlyMetSerVal	1286
DB	18524	-----AGTGAAGCCCGCGCTCACGACGCGGT-----	1849
QY	1286	IThrGlnCysSerLysGluAspGlyArgSerSerSerGlyPro-----ProH	1302
DB	18497	-----CGTAAAGCCCGGACAGGTCTCACTCTCGCGCGCGCG	1846
QY	1302	sgLutThrAlaAlaProLysArgThrTyrArgPheMetMetGlnGlyArgValGlyArgAla	1322
DB	18459	ACAGGCGCACGTGCACAGCTGTTCAGAGCGGACAGGAGTCCGGGGGGGTCTGCTCAACA	1840
QY	1322	eser-----SerAlaSerIleGlnGlyLeuMetGlyArgAlaIlePr	1336
DB	18399	CTCCCGTCCGCTGTGCACCAACGAGCGCGGTCTTCGGGGGCTCTGCAGGCTGCTTACA	1834
QY	1336	oProGlu-----ArgHisSerProHisIleLeuLysGlu-----	1347
DB	18339	GCCCCACCTGCGCGCGCGCGCGGTCTCGGCGGTCTCGATCATCACTGCAGCGCGCGCG	1828
QY	1348	-----GlnHisHisIleArgGlySer-----IleThrG	1357
DB	18279	CGCCCTCGCGCAGACACAGCGGCTCGGCGACCTGACCTCCGACGACCGGCTGCCCA	1822
QY	1357	ngLIleProArgSerTyrValGlnAlaGlnGlnLysArgTyrLeuArgArgGlu-----Al	1375
DB	18219	CCGGGCGCGCGCGCGCGACGACCATCTCCAGGTCCCGGTGCCGGAAGACACCATCTGGC	1816
QY	1375	AluValLeuLeuLys-----ArgGlnGlyThrProPro-----	1385

```

Db      18159 CAAACACCTCATGATCCGAGAGCCAGAGATGCTCCCGCAGCAGCAGCCCGCTGTGAA 18100
Qy      1386 -Pro-ProProSerSerArgAspLeuThiGluAlaTyrlsThr-----GlnAla 1402
Db      18099 GGTGCTCCCTCGCGCGGAGAGCTGTGCGGCGCGGAGAGAGCGGGTGTCTGCTGCT 18040
Qy      1402 euGlyProLeuLysLeu-Lys-----ProAlaHisGluGlyLeu 1414
Db      18039 TAGAGCCCGGCGAGCTCAGTGCTACGGCGCTTCGAGATATCAGCCAGTAAAGCTGGC 17980
Qy      1415 ValAlaThrValLysGluAlaGlyArgSerLeHisGluIleProArgGluGluLeu 1434
Db      17979 GCTGAACCGCTAGCTCGGAGCT-----CCAGAGACAGCAGC---CGC 17941
Qy      1435 HisThrProGluLeuProLeuAlaProArgProLeuLysGluLysLeuIleThrGlnGly 1454
Db      17940 CCGTGGCCCAAGACAGCCCGGTCCAGTCCACCT-----CGT 17905
Qy      1455 ThrProLeuLysTyzAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArg 1474
Db      17904 GCCCGTCAGCGTCCAGCTGCCAGCGCTCTGTACAGCTGCCAGAACCCAC---CTTCGT 17848
Qy      1475 SerLeuIleGlySerProGlyArgThrRheProProValHisProLeuAspValMetAla 1494
Db      17847 CCGCGTCAGCGCTCCACACCA-----CCCGCTCGCGCTCCCGCAGCGCGTGTGA 17794
Qy      1495 AspAlaArgAlaLeuGluArgAla---CysTyrgluLysLeuSerLysSerArgProGly 1513
Db      17793 GCGGCAATGCCAGACAGCGGGTGGCGCTACCTCCAGAACAGACAGCTGCCCTCTCCA 17734
Qy      1514 ThrAlaSerSerSerGlySerIleAla-Arg-----GlyAlaProVal11 1529
Db      17733 GAGGCTTCAGACAGCGCTCGGTCCAGGCGAGCAGCTCCGCGGAGGTTGGCGACAGTACT 17674
Qy      1529 eValProGluLeuGlyLysProArgGln-SerProLeuThrTyrgluAspHisGly- 1547
Db      17673 CCGCGTCAGCGCTCGCGCGAGACCTCTCTCAGCCTGTCAAGTAAAGCGAGCT 17614
Qy      1548 -----AlaProPheAlaGlyHis- 1553
Db      17613 GCGTCGCTTCGCGCTCAGAGCAGACAGCTCGCTCTTACTGGGAGAGCGGCTTCA 17554
Qy      1554 -----LeuProArgGlySerProValThrMetArgGluProThrProArgLeu 1570
Db      17553 TGGGCGCGGTGCCAGCGGTAGTGCATGTCATCTTCGCGAGAACACACCTTCGCG 17494
Qy      1570 lngluGlySerLeuSerSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrPro 1590
Db      17493 TCAGCTCCACATCAGCCCGTCCACCGCTCCACGTCCAGCCGACACCCAGT-CGAGCTC 17435
Qy      1590 rgGluIleAlaLysSer----- 1595
Db      17434 GACGTTCACAGCGCGCTTCGCAAGCGCTCCCGTACGGCGGATGCCGCTCTGCACC 17375
Qy      1596 --ProHisSerThrValProGluHis-----ProHisPro1 1608
Db      17374 TCGGAACCGCGCGCGTGCAGACAGCATCATGCCCCATCCCGACGCGTCCGACGGCC 17315
Qy      1608 lserProTyrgluHisLeuLeuArgGlyVal-----SerGlyValAspLeu 1624
Db      17314 TGGCTTCGAGCGCGCTACCCCGCTCTCCGCAAGCAGCAGCGCTCCGACACACC 17255
Qy      1624 yArgSerHisIleProLeuAlaPheAspPro-----T 1635
Db      17254 GCGCGGACACCTCGCGCTGAGCTGT-GGCCACCAACCGCGCAGGCTCCAGCCCGCA 17196
Qy      1635 hSerIleProArgGlyIleProLeuAspAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1655
Db      17195 CCGCCACCGCGCGCGAGACCCAGCAACCGAGAACAGCGGCGGTGCACCGTCCAC 17136
Qy      1655 euAlaProAsnProThrTyrlPro-----HisLeuTyrlProProTyrlLeuIleArgGlyT 1673

```

```

Db      17135 CCGCTCCAGCAGCGCTCTCTCCCGGCATCTCCGCGCAGACCCAGACAGACCA 17076
Qy      1673 yProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrlIleThrSerG 1693
Db      17075 GCCGTCCACGGCGCGAGCGCTCATCCGACCGCTGCA----- 17038
Qy      1693 lnglMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyL 1713
Db      17037 -----CCGCTCCGCAAGCCCGCGCTCTCTCCAGCAGCGCGCGCC 16995
Qy      1713 euSerProArgGluSerSerLeuAlaLeuAsnTyrlAlaIleArgGlyIleIleAla 1733
Db      16994 CATCCCGGCGCACTGGCTCTCT-----GCCGGGAAACAGAACACACCTT 16947
Qy      1733 sPLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAlaT 1753
Db      16946 GCTTCGCGCAGCGCGCTGATCCGCG-----ACACCGCGGATGGCGCGACCTTGCA 16893
Qy      1753 hTrAlaMetAspArgLeuAlaTyrlLeuProThrAlaProGlnProPheSerSerArgHis 1773
Db      16892 CAGCG-----CCCGCAGCGCTCTCTCGCTCCGACAGCTGCG 16854
Qy      1773 eSerSerProLeu-----SerProGlyLysProThrHisLeuThrL 1787
Db      16853 CCGAAGCACCGATGGCGCGTGCAGGAGCGAAGTGGCTCGGTACAGCGCGCTCTCTCAAC 16794
Qy      1787 yProThrThrThrSerSerSerGluArgGluArgAspArgAlaArgAlaArgAla 1807
Db      16793 GTCGACACCATCGACTCCCGGTCTCCGAGCAGCTTCGCGCAGCGCTCCGCGCTGCG 16734
Qy      1807 sPArgGluArgGlyLysSerIleLeuThrSerThrThrThrValGlnHisAlaProIleT 1827
Db      16733 CCCCACGAGGCGTGTCTCGCGCCGACAGAACGCGGAGATCGAACACCTCTCGTCC 16674
Qy      1827 rPArgProGlyThrGluGlnSerSerGlySerSerGlySerGlyGlyGlyGly- 1846
Db      16673 C-AAGGCTCGGCAGTGGCTCTCAACGCTTCGCGCGCGCTCTCAACGCGCTCTTCA 16615
Qy      1847 -----SerSerArgProAlaSerHisSerHisAlaHisGlnHisSerPro- 1862
Db      16614 GATGACATGGCGGT-CTTCGCGGTATCCGAAAGCAGACAGCCCGCGCGCGCGC 16556
Qy      1863 -----IleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeu-His 1879
Db      16555 CCGCGCTTCGCGCGCGCAGCGCGCTCTTCAGAGAACAGACCGCTCCCTCCAC 16496
Qy      1880 ---AsnThrGlyMetLysGlyLleIleThrAlaValGluProSerLysProThrValLeu 1898
Db      16495 CCAATGTCGGGCTCGGCTCTCCGCTGCAAGGTCGCGGACGACCTGTGCTGATC 16436
Qy      1899 ArgSerThrSerThrSerSerProValArgProAlaAlaAlaThrPheProProAlaThrHis 1918
Db      16435 GACAGCACCATGTAGACACAGCGGAGCGCGCGCA-----GCCCGCGG- 16388
Qy      1919 CyProLeuGlyGlyThrLeuAspGlyValTyrlProThrLeuMetGluProValLeuLeu 1938
Db      16387 TGTCCA---GGTTCGACTTCGAGCAGGCAAGGTACAGCGCGCTCGGCTTGGCGCCAG 16331
Qy      1939 ProLysGluAlaProArgValAlaArgProGluArg----- 1950
Db      16330 GCCCGAACCTTCACAGAGCGTCCGAGCTTCATGAGTGGCTCGCCAGTTCGCGCGTCC 16271
Qy      1951 -----ProArgAlaAspThrGly-----HisAlaPheLeu 1960
Db      16270 CATTGGCTTCACCGCGTGCATGTCTCGGGGAGACGGCGGACAGCAGAGAGCGCTGT 16211
Qy      1961 AlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGlu 1980
Db      16210 GGATCACCGCTGCTGGGAGGAGCGTTCGAGCGCGTCAACCTTGCTGGCAGCGTCTCT 16151
Qy      1981 ProArgProLeuVal-----ProProValSerGlyHis----- 1991
Db      16150 GGTTCACCGCTACCCACAGATCATCGCGCAGACAGCTACCGTGCAGCGCGCGCTCG 16091

```


ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/372,652
FILING DATE: 13-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/246001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ. ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1922 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-372-652-9

Alignment Scores:

Pred. No.: 1,08e-24 Length: 1922
Score: 718.00 Matches: 212
Percent Similarity: 48.81% Conservative: 76
Best Local Similarity: 35.93% Mismatches: 214
Query Match: 5.43% Indels: 88
Gaps: 25

US-09-522-753-5 (1-2517) x US-08-372-652-9 (1-1922)

QY 1975 ProserLysGlySerLysProAlaProLeuVal----- 1985
DB 209 CCAAGTGGAGGCCCGACCTCATGCTCAGTAGTATTCTGAGCGTGGAAAGATAA 268
QY 1986 -----ProProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsn----- 2001
DB 269 GGGCCCTCTCCAAATTCAGATATGAGGAAGACTAAGACCCGAGGAAAGACTACATT 328
QY 2002 LeuAlaProHisAlaSerProAspProProAlaProProAlaSerAlaSerAspPro 2021
DB 329 ACTGACGACTAATTCTATAGACGATACACCGCAATTGCTCGGACAAAGATGCG 388
QY 2022 HisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGlnLysLeu 2041
DB 389 AGGGAACGTGGCTCTCAAGATTCAAGCTCTCT-----AGTACCTTG 430
QY 2042 GlyTyrHisGlySerSerTyrSerProGlnGlyValGluProValSerProValSer 2061
DB 431 TCTTCTCAGAGTATATAACGGCTATGATGCTATGAGGTATAGTCCCGCAGCTCA 490
QY 2062 ProSerLeuThrHisAspLysGlyLeuProLysHisLeuGlnGluLeuAspLysSerHis 2081
DB 491 CCTGCACCAACCCGAGAAAGACACAGGCTATCAGCCAGACATGTTAAGCAANTCAA 550
QY 2082 LeuGlnGlyLysLeuArgProLysGlnProGlyProValLysLeuGlnGlyValAla 2101
DB 551 GGAAGAAATGAGTCCACTCGACAGATATGAAGTCCA----- 586
QY 2102 HisLeuProHisLeuArg-----ProLeuProGlnSerGlnProSerSer 2116
DB 587 ---CTGCATCATTTATGGTCCCGACGAGATCACCTTCCACAGCAAGCCACACTG 643
QY 2117 SerProLeuLeuGlnThr-----AlaProGlyValLysGlyHisGlnArgValValThr 2134
DB 644 CCCCCTCTTCCAGTCAAGGAGGATGGAGCAGGTGCCAGGACCCATGACTGATCACA 703
QY 2135 LeuAlaGlnHisIleSerGlnValIleThrGlnAspTyrThrArgHisIlePheGlnGln 2154

DB 704 CTTCGTGACCAACATCTGTCAATTAATACACAAAGATTGCTAGAAAT-----CAAGTT 757
QY 2155 LeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAla-----SerCysPro 2172
DB 758 CCCTCGACGACCTTCACTTCAATTCATTCACAACTCCACATCTGCTTGTGCATCACACT 817
QY 2173 ValLeuAspLeuArgArgProPheSerAspLeuTyrLeuProProProAsp----- 2189
DB 818 GTA-----AGAATAAAACCTCAAGCCGCTACAGCCAGAAATCAAGTCTCAGACT 868
QY 2190 -----HisGlyAlaProAla---ArgGlySerProHis-----SerGlu 2201
DB 869 GTCTTGCAATCCAGACACAGGCTCTTGAAGTCTTCCAGAAATCTGTGATTAATCCCG 928
QY 2202 GlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyLysAspGly 2221
DB 929 GGAAGACAGCCCGAAATATCTCCAGAGAGACTCATATC-----CATCAGAGCCC 979
QY 2222 IleGluProValSerProProGlnGlyMetThrGluProGlyHisSerArgSerAlaVal 2241
DB 980 TATGAGCCCATCTCCCAACCCCAAGCCCTGCTGTG-----CATGAGACGAGACAGC 1033
QY 2242 TyrProLeuLeuTyrArgAspGlyGlnGlnThrGluProSerArgMetGlySerLysSer 2261
DB 1034 ATGTGCTCTTGTCTCAGAGGGAAGTGAACCTGCTGACAAAGAGATTCATTCGATCA 1093
QY 2262 ProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGlnSerAsnSerAla 2281
DB 1094 CCAGAGAGATATAGCTATCTGCTCATCTTCCACCAAGCTT---GAAAGACATACACC 1150
QY 2282 MetValLysSerLysGlnGlnLysLeuLysLysLeuAsnThrHisAsnArgAsnGlu 2301
DB 1151 ATGGTTAAATCAAAACAGAAATTTTTCGTAAATTGAACCTCTGCTGAGAGTGAG 1210
QY 2302 ProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGly 2321
DB 1211 TCTGATATGACAGCTGCTCAGCAGAAACAGATCTTCAATCTGCTCAGCACTTACACA 1270
QY 2322 ThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGlnHisAlaSerThrAsnMetGly 2341
DB 1271 TCAGGTGACGATGAGCTCAAGAGCCATCTTTGTGATCCGCCAGT---AACCTGGGT 1327
QY 2342 LeuGlnAlaIleIleArgLysAlaLeuMetGlyLysTyrArgGlnTrpGlnLys----- 2359
DB 1328 CTAGAAGACATCATGAGAAAGCTCTCATGGAACTTTGATGATTAAGATCAT 1387
QY 2360 -----SerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAla 2375
DB 1388 GGTGTTGTCATCTCCATCTCTGTG-----GGCATTTATGCTGTGATGTCAGACACC 1438
QY 1439 TCAGTG-----GTGACGAGCAGCAGGACCGAGAGATGAAGGAGCA 1483
QY 2396 SerProGlyGlyGly---GlyLysAlaLysValSerGlyArgProSerSerArgLysAla 2414
DB 1484 TCACCTCATGACGAGGTATGCAACCAAGCGATCAACCAATCAACAGCAGGAAGTCT 1543
QY 2415 LysSerProAlaProGlyLeuAla-----SerGlyAspArgProProSerValSerSer 2432
DB 1544 AATATCTCCATTCCTGGGCAAACTATTAGAACTGAAAGGCTTCTTCTTCTCTCT 1603
QY 2433 ValHisSerGlnGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAsp 2452
DB 1604 GTGCATTCAGAGAGATTAACACAGCAGACACA-----GATGGCATGGGAAGAT 1657
QY 2453 ArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuLysMetArgLeuGln 2472
DB 1658 CGGCCCTCTTCAACAGGTTCTACTCAGTCCCTTCAACCCCTGACCTAGCAGGAGT--- 1714
QY 2473 AlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeu--- 2491

Db 1715 -----CTCAGACGTACACCACTACAGATCGCATGCCCATCTGCATCACC 1765
Qy 2492 ---AlaGlyProHis-----HisAlaTrpAspGluGlyProLeuLeuGly 2507
Db 1766 CAGACAGCTCCACATCAACAGAACCCGATCTGGAGAGGGAGCGCTGCCCTCTCA 1825
Qy 2508 SerGlnTrpGlnTrpLeuSerAspSerGlu 2517
Db 1826 GCGCAGTATGAGACATGTCTGATGATGAC 1855

RESULT 5
PCT-US95-16311-9
Sequence 9, Application PC/TUS9516311
GENERAL INFORMATION:
APPLICANT: Moore, David
APPLICANT: Seol, Wongi
APPLICANT: Choi, Hwang-Sik
TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING
TITLE OF INVENTION: POLYPEPTIDES AND RELATED MOLECULES AND METHODS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street, Suite 3100
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16311
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/372,652
FILING DATE: 13-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/246001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1922 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
PCT-US95-16311-9

Alignment Scores:
Pred. No.: 1,08e-24 Length: 1922
Score: 718.00 Matches: 212
Percent Similarity: 48.81% Conservative: 76
Best Local Similarity: 35.93% Mismatches: 214
Query Match: 5.43% Indels: 88
Gaps: 25

US-09-522-753-5 (1-2517) x PCT-US95-16311-9 (1-1922)

Qy 1975 ProSerIysGlySerGluProArgProLeuVal----- 1985
Db 209 CCAAGTGGCAAGGCGCCGCTCATGCTCAGTAGTATTCTGAGGCTGGAAAGATAAA 268
Qy 1986 -----ProProValSerGlyHisAlaThrIleAlaArgThrProAlaIysAsn----- 2001
Db 269 GGGCTCTCTCAAAATTCAGATATGAGGAAGAGCTAAGACCCGAGGAAAGACTACCAT 328

Qy 2002 LeuAlaProHisHisAlaSerProAspProProAlaProAlaSerAlaSerAspPro 2021
Db 329 ACTGACAGCTAACTTCAATAGACGATGACATACCCGGCAAAATGCTCTGCACAGAGATCG 388
Qy 2022 HisArgGlyHisThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeu 2041
Db 389 AGGAAACGTGCTCTCAAAAGTTCCAGACTCTTCT-----AGTAGCTTG 430
Qy 2042 GlyThrHisGlySerSerIysProGluGlyValAlaGluProValSerProValSerSer 2061
Db 431 TCTTCTCAGAGTATGAAGCGCTAGTATGATCCATTAAGATGATGATCCCGCAGCTCA 490
Qy 2062 ProSerLeuThrHisAspLysGlyLeuProLysHisIleGluGluLeuAspLysSerHis 2081
Db 491 CTGACACCAACCCGAGAAAGCCACAGCGCTATACAGCAGACATGCTTAAGCAAAATCA 550
Qy 2082 LeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGlyAlaAla 2101
Db 551 GCAGAAATGATGCCACTCGACAGATATGAGGTCCA----- 586
Qy 2102 HisLeuProHisLeuArg-----ProLeuProGluSerGlnProSerSer 2116
Db 587 ---CTGCATATTATGCTGCTCCAGCAGAAATCACTTCACAGAACCCACCATG 643
Qy 2117 SerProLeuLeuGlnThr-----AlaProGlyValLysGlnHisGlnArgValIleThr 2134
Db 644 CCCCACATCTTCCAGATCAGAGGGAATGGGACAGTGGCCCGACGCCATCAGATGATCA 703
Qy 2135 LeuAlaGlnHisIleSerGluValIleThrGlnAspTrpThrArgHisIleProGlnGln 2154
Db 704 CTGCTGACCAACATCTGTCAAAATATACACAGATATTTCTGAAAT-----CAAGTT 757
Qy 2155 LeuSerAlaProLeuProAlaProLeuTrpSerPheProGlyVala-----SerCysPro 2172
Db 758 CCTCGAGCGCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCA 817
Qy 2173 ValLeuAspLeuArgArgProProSerAspLeuTrpLeuProProProAsp----- 2189
Db 818 GTA-----AGAACTAAACCTCAAGCGGTTACAGCCCAAAATCACAGTCTCAGACT 868
Qy 2190 -----HisGlyAlaProAla---ArgGlySerProHis-----SerGlu 2201
Db 869 GTCTTCATCCAGACAGACAGCTCTAAGTCTCCAGAAATCTTGATGATAATCCCG 928
Qy 2202 GlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGluAspGly 2221
Db 929 GGAAGCAGCGCTGGAATAATCTCCAGAGAGGATCATATC-----CATGAAAGCCCC 979
Qy 2222 IleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaVal 2241
Db 980 TATGAGCCCATCTCCACACCCCAAGGCGCTGTGTG-----CATGAAAGCAGACAGC 1033
Qy 2242 TyrProLeuLeuTrpArgAspGlyGluGlnThrGluProSerArgMetGlyLysSer 2261
Db 1034 ATGTGCTCTTTCACAGAGGAGGTGACCCCTGTCGACAAAGAGATGATTCATCTCA 1093
Qy 2262 ProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuLeuThrGluSerAsnSerAla 2281
Db 1094 CCAAGAGATATGACTTCTGCTTCACTTCTTCAACCAAGTT--GAAAGCAGATCAACC 1150
Qy 2282 MetValLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGlu 2301
Db 1151 ATGCTTAATCAAGAAACAGAAATTTTTCGTAATGAACTCTTGATGAGAGTAC 1210
Qy 2302 ProGluTrpAsnLysSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGly 2321
Db 1211 TCTGATATGACAGCTCTCAGCAGGAGACAGATCTTAATCTGCAGAGATTAACACA 1270
Qy 2322 ThrGluLeuMetThrTrpArgSerGlnAlaValGlnGlnHisAlaSerThrMetGly 2341
Db 1271 TCAAGTCAAGTGAAGCCAAAGAACCACTTTCTTGTGATGCCGCCAGT---AACCTTGG 1327
Qy 2342 LeuGluAlaIleIleArgLysAlaLeuMetGlyLysTrpAspGlnTrpGluGlu----- 2359

```
|||||
Db 1328 CTAGAAACATCATCAGAAAGCTCTCATGGAGAGTTTGATGATPAAAGTTGAAGATCAT 1387
QY 2360 -----SerProbleuSerAlaAsnAlaPheAsnProleuAsnAlaSerAla 2375
Db 1388 GGTGTTGTCATGCTTCATCTGTG-----GGCATTATGCTGCTGTGTGTGCGACACC 1438
QY 2376 SerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThr 2395
Db 1439 TCAGTG-----GTACGACACGACGACGACGACGAGATGAAGGGAGCCA 1483
QY 2396 SerProGlyGlyGly--GlyIysAlaIysValSerGlyArgProSerSerArgIysAla 2414
Db 1484 TCACCTCAGCAGAGATAGCAAAACCAAGCTGATCAACAATCAAAACAGCAGAACTCT 1543
QY 2415 LysSerProAlaProGlyLeuAla-----SerGlyAspArgProProSerValSerSer 2432
Db 1544 AATCTCTCATTTCTGGGCAAGCTATTGAGAACTGAAAGGCTTCTTCTCTCTCTCTCT 1603
QY 2433 ValHisSerGlyGlyAspCysAsnArgArgThrProLeuThrAsnArgValTropGluAsp 2452
Db 1604 GTGCATTCAAGAGATATACCAACAGCAGACCA-----CGATGGGCAATGGGAAAGT 1657
QY 2453 ArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGln 2472
Db 1658 CGGCTCTTCAACAAGGTTCTACTACTGTTCCCTTACACCTCTGACCATACGAGATG--- 1714
QY 2473 AlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeu--- 2491
Db 1715 -----CTCAGCAGTACACACCTTACACAGATGCATGCCGCCCATCTGCCATCACC 1765
QY 2492 ---AlaGlyProHis-----HisAlaTropAspGluGluProIysProleuLeuCys 2507
Db 1766 CAGACGCTCCATCATCAACAGAACCGCATCTGGGAGAGAGGAGGCTGCCGCTCTCTCA 1825
QY 2508 SerGlnTyrGluThrLeuSerAspSerGlu 2517
Db 1826 GCGCAGTATGAGACACTGTCTGATAGTGAC 1855

RESULT 6
US-09-902-540-1090/c
/ Sequence 1090, Application US/09902540
/ Patent No. 6833447
/ GENERAL INFORMATION:
/ APPLICANT: Goldman, Barry S.
/ APPLICANT: Hinkle, Gregory J.
/ APPLICANT: Slater, Steven C.
/ APPLICANT: Wiegand, Roger C.
/ TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
/ FILE REFERENCE: 38-10(15849)B
/ CURRENT APPLICATION NUMBER: US/09/902,540
/ PRIOR FILING DATE: 2001-07-10
/ PRIOR APPLICATION NUMBER: 60/217,883
/ NUMBER OF SEQ ID NOS: 16825
/ SEQ ID NO 1090
/ LENGTH: 14462
/ TYPE: DNA
/ ORGANISM: Myxococcus xanthus
US-09-902-540-1090

Alignment Scores:
Pred. No.: 7,98e-23 Length: 14462
Score: 698.00 Matches: 652
Percent Similarity: 32.50% Conservative: 301
Best Local Similarity: 22.24% Mismatches: 1028
Query Match: 5.28% Indels: 964
DB: 4 Gaps: 143

US-09-522-753-5 (1-2517) x US-09-902-540-1090 (1-14462)
QY 44 HisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGlnProGlnArg 63
|||||
|||||
|||||
```

```
Db 14437 CACCCAGCCGCTCCAGCGGGTAGTCCGCTCCAGCGGACGACATACCGCCCCGCTTC 14378
QY 64 ArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSer-GlnGluLeuHis 83
Db 14377 AGAATCCCAAGCGGCGCTCACCACCCAC-----TCCAAAGAGCGCTCCGCGACATATCC 14324
QY 83 LeuArgProGluSerHisSerTyrLeuProGluLeuGlyIysSerGluMetGluPhe11 103
Db 14323 ACTCCACCTCCGGGCG----- 14308
QY 103 eGluSerLysArgProArgLeuGluLeuLeuProAspProleuLeuArgProSerProle 123
Db 14307 -----CCACTCCCAT 14297
QY 123 uLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrG1 143
Db 14296 CCCCTCAGGTGTGGGCGCAACTGGTTTGCTGGCATTCACAGCCCTCGTATCACCCTGC 14237
QY 143 LysLeuGluProValSerProProSerProProHisThrAspProGluLeuGluLeuVa 163
Db 14236 TGCCTGCCACTCCACTGCCACCGCGCGGGGCTCTCCACTGCCTTCAACAC 14177
QY 163 LProPro-ArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgGlu- 182
Db 14176 TCCGCCAGCGA-----CGCTCCCTCGGGTACTCCGCTCCG 14141
QY 183 -----IleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnLeuG 201
Db 14140 CTCGGCTGCATCTCCACCGACGAGCGGCGGCTCTCCACCGACATCAACGACAGCTCG 14081
QY 201 LngGluGluAlaAlaLysProProGluProGluLysProValSerPro----- 216
Db 14080 CCCAGAGC-GCGTCTCGGATTGCGACCGCGGCTCCAGACACCGACATGCTGCTCAG 14022
QY 217 -----ProProIleGluSerLysHisArgSerLeuValGlnIleTyrAspGluA 234
Db 14021 CAGCGCGCTCCGCGCTTC-GCCTGAAACAGTCCGCTGTGATATG----- 13976
QY 234 snArgLysLysAlaGluAlaAlaHisArgGlyLeuGlnGlyLeuGlyProGlnValGlu 254
Db 13975 -----ACTGACGCGGGGCGAC- TTCGGGCACTCGACGACCAACATCGACAGTGTGAAC 13922
QY 254 euProLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnG 274
Db 13921 TTCGACGTCTGAATCCG-----CGGCTCATTC 13895
QY 274 InAlaMetArgLysLysLeuIle-----LeuTyrPheLysArgArgAsnHisAlaArgLysG 293
Db 13894 ACCGTCAGCGAACAATCCTGTGCGACCTTCACTTCGGGACACAGCGCTGTGGAATGTC 13835
QY 293 In-----TrrLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGlu 311
Db 13834 AGGAGACTTGAAGAACAGCGGCTGGGACATCAGATCCCGCTCGGCTGCAATCCCTCAC 13775
QY 311 yAlaValAlaArgIleGlu-----AsnAsnProA 321
Db 13774 AGCTTCTCAACGCGGACGTCCTGTGCGGTAAGGCGCCAGGAGTCACTCACGACATGTC 13715
QY 321 rGArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleA 341
Db 13714 GCCAACATGTCCCGGAAGGAAGCTCGGACGTACGACAGCCCTCAGCACACAGGATATG 13655
QY 341 rGlySerGlnArgGluLeuGlnGluArgMet-----GlnSerArgValGlyGlnArgG 358
Db 13654 GCGAAGTTCCGATGAGCCCTCCGCTCGGCTGCGTGCAGACCCGCGATGGGAAGCCC 13595
QY 358 LysSerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspG 378
Db 13594 ACGCTGATGTGCTCCGCGCCGACGAGCCGCAACACACGCTCG----- 13551
QY 378 LysLeuSerGluGlnGluAsnLeuGluLysGlu-----MetArgGlnLeuAlaV 394
Db 13550 --ATACGCGCGCAAGACACATGAACGGCGTGGCCCTTTCGCGCGCGGCTGCCTCC 13493
```

QY 394 alleProPomeLeuTyzAspAlaapGlnIArgIleLyPheIleasnMetAsnG 414
::: |||
Db 13492 ACTTGCGCGCA-----CAGGCCCTTCGG----- 13470
QY 414 lyeuMetAlaapPromeLlyValTyryIyAspArgGlnValMetAsnMetTrpSerg 434
13470 ----- 13470
QY 434 lngIngluLySgluThrPheArgGluLyPheMetGlnHlspProLyAsnPhenLyIeu 454
13469 -----CCACAT----- 13464
QY 454 leaIaSerPheLeuGluArgLyThrValAlaGluCyValLeuTyTyryIeuThrL 474
13463 -----CTGC 13460
QY 474 yelyAsnGluAenTyryLySserLeuValArgSergTyryArgArgGlyLySserG 494
::: |||
Db 13459 CGCTGATTCACCTCGAAGAGTCTCGCGCGGAGGAGCCCTGTCGCGGAGATCC 13400
QY 494 lngIngluIngluIngluIngluIn----- 502
Db 13399 AGSAATCGCGCGCGCCGCCAGCACTCGCGCCAGTAGCAACCTGCGATTCCAAACACC 13340
QY 503 -----GlnGlnG 505
Db 13339 TCGTCCTTCAACACTCGCGCTCGCAGAGCGCGTACGCGCGTACGCGGCAAGCTCC 13280
QY 505 lngIngluIngluIngluInPromeCProArgSergGlnGluLyAspGluLySgluL 525
::: |||
Db 13279 GGGAAATGTCGACCGGACCTTGG---CGAAGACCGCTGTAAGCGCGCGCAACTCCGA 13223
QY 525 ysgLluLySgluAlaGluLySgluGluLySProGluValAlaGluAsnAspLyGluAspL 545
::: |||
Db 13222 AGCAGACAGCTCGCGGACCAACCTCGCAGACAAATGTGATGCAACAGCAGCAGAGAGCG 13163
QY 545 euleuLySgluLyThrAspAspThrSergLyGluAsnAspGluLyGluAlaValA 565
::: |||
Db 13162 TGGG-CCTGCTCGGACTGCTTCAACAGCAGCG----- 13132
QY 565 laSerLySgluLyArgLyThrAlaAsnSergGlnLyArgLySgluLyArgLyS 585
::: |||
Db 13131 -----GGCGCAGACGCGCGCCACGCGCAGACTCGAATGCGCGGCTCCCGCGCG 13080
QY 585 etMetAlaAsnGluAlaAsnSergGluGluAlaIleThrProGlnGlnSeraAlaGluLeuA 605
::: |||
Db 13079 CA---CCCGGTCCCGACCC-TCCGACTCATCTCGGCGCGCCAGCTCCACACGAGAGGT 13024
QY 605 laSerMetGluLeuAsnGluSeraArgTrpThrGlnGlnGluMetGluThrAlaVal 625
::: |||
Db 13023 GAGACCGCGCAT---GGCGAAGCAGATCGCTCGACAGGTGCTCCGCTTCACTCCGGAAG- 12968
QY 625 ysgLyLeuLeuGlnHlsgLyArgAsnTrpSeraIleAlaArgMetValGlySeraTr 645
12967 -----GCCGTACGACGAGACTTCACTGACGCGCA 12940
Db 645 hrValSergGlnCyLySAsnPhenTyryIySAspArgGlnAsnLeuAsp---- 663
12939 CCAAGCTCCAG-----AAGGCTCTCTCAAGGCAAGAAACAT 12904
QY 664 -----GluIleGlnGln-HlslYsLeuLyMetGluLySgluArgAsnAlaArg 680
::: |||
Db 12903 CCAGGGGCGCTCCACCGCAGACACATACGGTCCGTTGACTCGGTAACCCCGCGCTTCA 12844
QY 681 ArgLyLyLyLyLyAlaProAlaAlaIaSerGlnGluAlaIa---PheProProVal 699
12843 GCCGATCCAGAACCAACAGTCTGCTGCGCGAATACAGCGGCAACCCGCTCCCGCGAG 12784
QY 700 ValGluAspGlnGluMetGluIaSerGlyValSergLyAsnGlnGlnGluMetValGlu 719
12783 GCATCGGACGACGCGGAGCAGCGCTGACTTC----- 12751

QY 720 GluAlaGluAlaLeuHlslAsnIaSerGlyAsnGluVal-----ProArgLyGlu 735
12750 GAGTCGCTCTCACTCGCTGGGCAAGAGCCGAAAGTGGGGCTTCGAAGGCGCG 12691
QY 736 CySergGlyProAlaThrValaAsnAsnSeraSAspThrGluSerIlePserProHis 755
12690 GTACCGGAGACTTACGC-----CCAGGACCCGAC 12661
QY 756 ThrGluAlaAlaLyAspThrGlyGlnAsnGlyProLySProProAlaThrLeuGlyAla 775
12660 GCACAGCGC-----CCACACTCGCTGCGCAATGAGCAAT 12625
QY 776 AspGlyPro-----ProProGlyProProThrProProArgTrpSera 790
12624 GACCACCCGACGGGAAGAGTCTCTCCCGGCCACCGCTCATCCGCAAG----- 12571
QY 791 ArgAlaProIleGlu-----ProThrProAlaSergL 801
Db 12570 CGTGGGCCGACAGACGGGCGCAACGACTTTCATGGGGAGAGCGCGCTCATAGCGG 12511
QY 802 AlaThrGly---AlaProThr-----ProPro----- 809
Db 12510 GACCGGCGCACGGCGCGACCTTGGCAGCGGACGCGCGCCAGTCCACTTGCCAT 12451
QY 810 -----ProAlaProPserProSeraIaPro----- 818
Db 12450 TGGCGCTCAGCGGCGGCGTCCAGCACCAAGCCGACACATGTCGCCAGAA 12391
QY 818 ----- 818
QY 12390 GCCGCTGCTGAGAAAGCCCTGAGGAGCGCGATGTCGCGAGCGCGCTGCTACGAA 12331
QY 819 -----ProProAl-----ValProLyS 824
Db 12330 CGCGTACGCCACACCGCATTTGTCACAGGACCTTCTCGCGGACCAACACCGCGT 12271
QY 825 GlnGluLySgluGlnGluThrAlaAlaIaProProValGlnGluGlnGlnLyS 844
::: |||
Db 12270 CCTGAGACGTGCGGATGCGCTGAGACACCGCGCCACTTC-CCGGCTTCACCGGAA 12212
QY 845 ProProAlaAlaGlnGluLeuAlaValaAspThrGlyLyAsnIaGlnGluProValLyS 864
12211 CCT-----CTCACCTTACCTGAGAACCTCGCGCGCGGAGAACTCC 12170
QY 865 Glu-----CyThrGlnGluAlaGlnGluGlyProAlaLySgluLySAspAlaGluAla 882
12169 AGTCGCCCATCCCCAACCAACGCGACCTTGTCCCGTGGGTAATCTCGCTCCAGCC 12110
QY 883 AlaGluAlaThrAlaGlnGluAlaLeuLyAlaGluLySgluGlySergLyArg 902
::: |||
Db 12109 TCCATGCTGACGAGGTCCGCGACGTACCGCTCGCGCTC-----AACTCGGGCGC 12059
QY 903 AlaThrThrAlaLySergSergLyAlaProGlnAspSeraSAspSeraIaThrCyS 922
12058 CATACGTACCTCGCGTACGCTGCGGCTCCACAGAACACTTCCCCCAGCGCCAC 11999
QY 923 AlaAspGluValaAspGluAlaGlnGlyGlyAspLyAsnArgLeuLeuSer-ProArgTr 942
11998 -----GGCACAGACGAGGCTCGGCTCCAGACAGCTAA 11966
QY 942 oSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLyS-----Pr 960
11965 ACCCGTGTGTGCCATTCGCGCGTCCA-----GGGCACTCTCGCGGCTGACTCC 11915
QY 960 oLeuAspLeuLySglnGlnGlnArgAlaAlaIleProProIleGlnVal----- 978
11914 GCTCGAGCG-----GCAGCTCACTCGCGCGCGCACACACCGCTC 11870
QY 979 -----ThrLySValHlsgLupProArgL 987
Db 11869 GTCTCGCTGCGCGCTGAGTGTGTCACCTCGCACCCGAGGCCACCGCTCCCTCG 11810
QY 987 uAspAlaAlaProThrLySProAlaProProAlaProPro----- 1001

```
Db 11809 TCACAGAGCTCCACGAGGCGG---CCTCTCTCCACACGACCGGCTTCGTCGCAAC 11753
Qy 1002 -----ProGlnLeuGln-----ProGlnLeu 1009
Db 11752 ACCTGCGCGGCGCGCGCCACTCATCAGCGCGGCGGAGTGCGAGCCACACCTTCANA 11693
Qy 1009 rAaPAlaProGlnGlnProGlySerSerProArGlyLySerArGserProAlaProPr 1029
Db 11692 CACTCCACCCCATGCGCGCTGGAAGTACTCTCCACTCCGCGCGCGCTGCTTGCCTCTCC 11633
Qy 1029 oAlaAspLySgLuAlaPheAlaAlaGlnAlaGlnLySLeuProGlyAspProProCysTr 1049
Db 11632 TCGGTCAGCAGCATGCA---GCAgTCCCGCCGTCACAGCGCGCGGTCAGACACCGGTGTT 11576
Qy 1049 pRhSerGlyLeuProPheProValProProArGlyValLLeLyAlaSerProHisAl 1069
Db 11575 C-----CCAGGTCCCGCGCACCGCTGCAGACAGCGCGGTCGTCGACCTCC 11528
Qy 1069 aProAspProSerAlaPheSerTyAlaProProGly-----ThrlLeSe 1081
Db 11527 GCCAACCCACGCGCGCATTCGCGCGCTTCACGTAAGGGAAGCTGTCGTCGAC 11468
Qy 1081 -----ThrlLeSe 1081
Db 11467 GCCAGCGCTTGCGCGCGCGCGCTTCGCGATGTGACAGACGTAAGCGCATGTCGTCG 11408
Qy 1082 ---HisProLeuProLeuGlyLLeuHisArGProAlaArgProAlaLeuProArg----- 1098
Db 11407 GGTTCGACCTCTAACCGCGGTGCTTCGTGCGCTGCGCTTCGAGCGCGCGGCTGTCCA 11348
Qy 1099 ---ProPro-----ThrlLeSe 1103
Db 11347 TCCATCTCCACCGCGCGCGGAGCAGTCCCTCGAAGTGGCGCATGCGCGCTGTCGTC 11288
Qy 1103 rAsnProPro-----ProLeuLLeSerSerAlaLySHisProSerValLeuGlnArgG 1121
Db 11287 ACCACACACGCGCGCGCGCATTCCTTCACACGCGCGCGAGCGGAGTCTG----- 11236
Qy 1121 nLLeGlyAlaLLeSerGlnGlyMeSerValGlnLeuHisValProTySerGlnHisAl 1141
Db 11235 -----GTTGCGAGGGTTCACGCGGACGTAAGCGCTCCCGCTTCAGAAATCCCC 11186
Qy 1141 aLySaLaPro-----ValGlyProValThrMeGlyLeuProLeuProMe 1156
Db 11185 AGCATTCGCCACGACCCACTTCACGCGCGCTCCAGGACACAGCGCCACTGCACTTCGGG 11126
Qy 1156 caSPrPolySlySLeuAlaProPheSerGlyValLySglnGly-----GlnLeuSe 1173
Db 11125 CCCACCC-----CCATTCCTCCCACTGGTGGCGCCACTGTTGGCTTCGCGC 11078
Qy 1173 rProArGlyGlnAlaGlyProProGlyLeuGlyValProThrlAlGlnGlnAlaSe 1193
Db 11077 TCCAACCTGCGGTAAGTCACTGCTCCGCTCCGACCGGACCGCGGCGCTTCGCGCTC 11018
Qy 1193 rValLeuArGlyLy-----ThrlAlaGlySerValProGlyLySerLLeThLySg 1211
Db 11017 CGCTCGCTGCTGCTGCTGCAACACCGGTGAGACGCTCCACCGGAGACGTGGGCGAG 10958
Qy 1211 LyLLePro-----SerThrlArgValProSerAspSerAlaLLeThTr 1225
Db 10957 CCGTTCATTCACGACGACCTGATGACGCTCCGATCGTCAAGC-----CGGT 10907
Qy 1225 yArGlySerLLeThrlSgLyThrlProAlaAspValLeuTyLySgLyThrlLLeThA 1245
Db 10906 ATCTGGGAGAGCGCGCATTCGGGCGCGGTGAGC-----GGCTTCACGACCC 10859
Qy 1245 rGllLLeGlyGlnAspSerProSerArGLeuAspArGlyArGlnAaP----- 1261
Db 10858 TGTCTCAGGTAGAACACATCCGCTCGC-----CGTGGCGCGCGTGAAGAAGTGGTG 10805
Qy 1262 -----SerLeuProLySgLyHisValLLeTyG 1271
```

```
Db 10804 CTGAACCTGAGACAGCCCTCGAAACCCCGGTGAGTCTCCCGGCTCAAGCAAGCTCG 10745
Qy 1271 LuGlyLyLeuGlyHisValLeuSerTyGlnGlyGlyMeSerValThrlGlnCysSerL 1291
Db 10744 AAGCGAGCGC-----CTCTTGTTCATCAACCCCATC 10712
Qy 1291 ySgLuAspGlyLy-----ArgSerSerSerGlyProProHis----- 1302
Db 10711 GCGTTCACGCTCAGGAGGAGGAGCGACGACAAAGCGCCACCGGCGCTTCGACACAGAA 10652
Qy 1303 -----GlnThrlAlaLLePro-----LyArGThrlTyAspM 1313
Db 10651 AGACCTTGAAAGAGCGCGCTGCGCGCCACAGTTCCTCCGCTGTCAGATCTTCACAGAC 10592
Qy 1313 eMeGlyGlyArGValGlyArGAlaLLeSerSerAlaSerLLeGlyLeuMeGlyA 1333
Db 10591 TTCTCGAAGGGAAGTTC-----CTGATG----- 10569
Qy 1333 rGAlaLLeProProGlyLLeThrlSgProHisProHisLLeuLySgGlnHis-----H 1350
Db 10568 ---CTGTCAGCGCCCGACGCT---GCGTTCACGACCTGCGCGCAGATTCGCGGAAGTG 10514
Qy 1350 LeLLeArg-----GlySerLLeThrlGlnGlyLLePro----- 1360
Db 10513 GCATGAGCTGGAACCTGGCGCGGAGAACCAAGGCTTTCAGAAACCAAGCATGAGCC 10454
Qy 1361 ---ArgSerTyValGlnAlaGlnGlnAaP-----TyLeuArgA 1373
Db 10453 TCCGCTCTGCTGTAACCGGAGACCGCGCATGGGCGAGCCACAGACGTCCTTCGCGCC 10394
Qy 1373 rGlnAlaLyLeuLeuLySArGlyGlyThrlPro-----P 1385
Db 10393 GAGTACCGGACAGCAGGAGCTGGAAGCGG---CGAGACGACATGAAGGCGTTC 10335
Qy 1385 roPro-----ProProProSerArGArPLeuThrlGlnAlaTyLLeThrlGlnAlaLeuG 1403
Db 10334 CTCCGATGAGGACAGTCCGATGACCGGCTGCTCATCGGCGCTGACACCGCACCGC----- 10279
Qy 1403 LyProLeuLyLeuLySProAlaHisGlyGlyLeuValAlaThrlValLySglnAlaGlyA 1423
Db 10278 -----GCACCCATGACACCGCGGT-----GCGAACAGTGCAGCGGCTGCGG 10239
Qy 1423 rGSerLLeHisGlnLLeProArGlyGlnGlyLeuAlaGlnHisThrlProGlnLeuProLeu--- 1441
Db 10238 CTGTCGCTGAGTACCTCCAGGTGAGAGGCGCGCAGACAGCTGCTGCGCCACCAAGTC 10179
Qy 1442 ---AlaProArGProLeuLySglnGlySer---LLeThrlGlnGlyThrlProLeuLyTyA 1460
Db 10178 GAGCTGCGCTTCACAGACGCGCGCGCTCAGCCATTCGCGCTCCACACCGCGTAAGTCGCG 10119
Qy 1460 ArPThrlGlyAlaSerThrlThrlGlySerLyLyHisArPValArGSerLeuLLeGlySerP 1480
Db 10118 GATCTGACAGGCGAGCTCCGCGACACCGGAGAACCGTCTTCGTA-----GGAAGGC 10068
Qy 1480 roGlyArGThrlPheProProValHisProLeuAspValMeKAlaArGAlaGlnAlaLeuG 1500
Db 10067 CCGATAGAGCGCGCCACCTCCCGGACCA-----GCAAGCCGAGGAGACAGCAATC 10017
Qy 1500 LuArGAlaCyEThrlGlnGlnLeu-----LySerArPProGlyThrlAlaSerSers 1518
Db 10016 CGAGAGATGTATGACAGACACGACACGAGGCGTGCAGCTCCGCGAGACCGCGACAG 9957
Qy 1518 er-----GlyGly-----SerL 1522
Db 9956 CCGCAGCGCACAGAGGCGCGCGCTCCAGATGAACCGCGCGAGGCGTTCGATCGCG 9897
Qy 1522 LeAlaArgGlyAlaProValLLeValPro-----G 1532
Db 9896 AGCGCGCGGAAATTCGAGCTCCGCTGCTCACTTCGCGAGCGGTCTCAAGTCTCTGAAGGC 9837
Qy 1532 LuLeuGlyLySPro-----ArgGlns 1539
Db 9836 GAGCTGAGCTCATAGCGCGGATGATGACCTGAAAGGCTGAGCTCTGTTGGAGGCAAA 9777
```


QY 1539 exProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerP 1559
Db 9776 CGTCGTGGCGA---ACGACTCATGGCGCGCCACC-ACCGCGGAAAGACCGCGCTGCACAG 9721
QY 1559 roValThrMetArgGluProThrProArgLeuGlnGlySerLeuSerSerLyAs 1579
Db 9720 CGCTCAATCGAAGGACCGCTTCGATCCGAG-GACGATGGGAGAGCTTGTAGCTGGCACTG 9662
QY 1579 laSerGlnAspArgGlyLeuThrSerThrProArgGluLeu-----AlaL 1594
Db 9661 CCGGAGATGGAG-----CTGTGATGATACACAGACCGCGCTGGGGAACAGACAGCGGC 9608
QY 1594 ysSerProHisSerThrValProGlnHisProHis----- 1606
Db 9607 AATGCGCGCATGGAGATACCGGGTCAGCGCTGCACAGCGCTCGCGCGCGCTTCCCGC 9548
QY 1607 -----ProLleSerProTyrGlnHis----- 1613
Db 9547 GCCGACTCCGTGATTCGACGCGCGCGCTTCGACAGATGACGGAAGCGCTGCGC 9488
QY 1614 -----LeuLeuArgGlyValSerGlyValAspLeuTyr-----ArgSerH 1627
Db 9487 AGGGGCACTCCACGCGGAAAGCGAGGATGTGCGCATGAGCTGCGCGCAAC 9428
QY 1627 laIleProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAla 1647
Db 9427 GAGTGGCGCGGCAACTCGAAGAGTCAGCGTC---CGCGCTCACCGCTGATGCC----- 9375
QY 1647 laAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProP 1667
Db 9374 -----CAGCACTTCACACCA----- 9360
QY 1667 roTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGlnAspArgGlnThrIleLea 1687
Db 9359 -----GATGATGCTCAGCTTCGTGCTCG 9338
QY 1687 snAspTyrIleThrSerGlnGlnMetHisAsnThrAlaThrAlaMetAlaGlnArgA 1707
Db 9337 GTGAC-----CGTACCGCGC 9323
QY 1707 laAspMetLeuArgGlyLeuSer--ProArg-----GluSerSerLeuAlaLeuA 1723
Db 9322 TCATATGCTGAGAGGGGCTCGGAAGCGCGCGTCGAGGACCAATGCTTGGCGTGC 9263
QY 1723 snTyrTrAlaArgIleProArgGlyIleIleLeuSerLeuGlnValProHisLeuProValL 1743
Db 9262 ACCTTCCCGTGGCGCTCAG-----CGGACGCGCTTCAGACCA-CCA----- 9220
QY 1743 euValProProThr---ProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuP 1762
Db 9219 -----CCACCGAGGAGCACTGATCTCGGCACTTCTGCTCACT-----ACTC 9174
QY 1762 roThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGlyGlyP 1782
Db 9173 CTTCACAGCGCGCTCTCTCTCTCCGCGGCAACAGTACGCT---CCACCGGCTTCTC 9117
QY 1782 roThrHisLeuThrIlyuProThrThrThrSerSerSerGlnArgGluArgAspArgAspA 1802
Db 9116 CTGCGCCCTCCCTCACCACCGCTGCG----- 9088
QY 1802 rgGluArgAspArgAspArgGluArgGluuysSerIleLeuThrSerThrThrThrValG 1822
Db 9087 -----CTTCTTCACTCGCGCGCTGCTG 9066
QY 1822 luHisAlaProIleTPrArgProGlyThrGlnInSerSerGlySerSerGlySerSerg 1842
Db 9065 CTTCAGCGCGCTCTCCACCTCCCAACTCA-----TCGGTACCCCTCACTTTCAC 9012
QY 1842 lyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHisSerP 1862
Db 9011 CTGCTGTCTCCAGCGCCCAAGAAATCTCAGACGACCACTCTCTGCACTCACTCGTCTC 8952

QY 1862 roLleSer-----ProArgThrGlnAspAlaLeuGlnGlnArgProS 1876
Db 8951 CCGCGTCCGTATACACCGCGTCCCGCTCCCAAGGAGAGGAGCAAGAGCGTCCGCT 8892
QY 1876 euValLeuHisAsnThrGlyMetLyGlyIleIleThrAlaValGluProSerLyAsProT 1896
Db 8891 CCACTCCGCGCTCC-----CCAGTACCCCTCCGCAAGCGCTC 8853
QY 1896 hrValLeuArgSerThrSerThrSerSerPro-----ValArgProA 1910
Db 8852 TCTCCCAAGTACAACTCTCCCTCCACCCCGGAGCGGACGATTCCTTGTCTCAG 8793
QY 1910 laAlaThr-----PheProAlaThrHisCys-----ProLeuG 1922
Db 8792 CAGTACACCGCGGTATTTCCACCGCTTCCCAATGGGACCGTCAACCCCACT----- 8737
QY 1922 lyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuProLyGlyA 1942
Db 8736 -----CTTCTCTGCGACCGCGGTGCACTGTGTAAGTCTGT 8700
QY 1942 laProArgValAlaArg-ProGluArgProArgAlaAspThrGlnHisAlaPheLeuAla 1961
Db 8699 CCGCTGTCTGCGCGCGTACCGCTTGAGAGAGATGCCCACTCTCACTCT---CTTCTCT 8643
QY 1962 lyAspProAlaArgSerGlyLeuGluProAlaSerSerProSerLyGlySerGluPro 1981
Db 8642 CACCGCGCGGACGACGACGCTCTCTC---CGCGACGACCTGCTCACTCACTGCGC 8584
QY 1982 ArgProLeuValProProValSerGlyHisAlaThrIle---AlaArgThrProAlaLyS 2000
Db 8583 CTTCGCGGTGTAGCGCTGCTGCTGCTGCGAAGAGGCTGCGTACAGACAGGACGTCA 8524
QY 2001 AsnLeuAlaProHisHisAlaSerProAspProProAlaProProAla---SerAlaSer 2019
Db 8523 CCGCTCTCTCTCAGCTCTCGCGCGCGCTTCCCACTTCCAGCGCGCTTCCGCGTACA 8464
QY 2020 AspProHisArgGlyuThrGlnSerIlyuProPheSerIleGlnGlnLeuGlnLeuArg 2039
Db 8463 CACACAGCTTCCCGCGCTGACGACGCGCGC----- 8431
QY 2040 SerLeuGlyTyrHisGlySerSerTyrSerProGlnGlyValGluProValSerProVal 2059
Db 8430 -----ATACCTCAACGTCGACGCGT 8410
QY 2060 SerSerProSerLeu-ThrHisAspLyAsGlyLeuProLyHisLeuGlnGlyLeuAspLy 2079
Db 8409 CGAAGCCGCTGCGCGCA-----TTGACGCCACACTCTCTCGGCGCGAGGT 8362
QY 2079 sSerHisLeuGlnGlyGluLeuAspPro-----LyAsGlnProGlyProVa 2094
Db 8361 CCGGACGCGCGCGCGGCAACACAGCGCGTTACCGGACGCTGCGGACCGGACCGCTCT 8302
QY 2094 lyuysLeuGlyGlyAlaAlaHisLeuProHisLeuArgProLeuProGlnSerGlnPr 2114
Db 8301 TCGGATTTCCCGGTCTTCCGAGTAAATACATGACGTAACCGGAGTGGCGCTCTCTAC-- 8244
QY 2114 oSerSerSerProLeuLeuGlnThrAlaProGly-----ValLy 2127
Db 8243 -TCGGGCTCCGAGTTGTCTGTGCTGCTGCGCAATGCGCTGCCCACTCGGTTCCCA 8185
QY 2127 vGlyHisGlnArgValAlaThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTy 2147
Db 8184 CGGACACACCAACTGCTTACCGGACGCACTTGTCCGCGCACTTCTCTGCGGACCA 8125
QY 2147 rThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheP 2167
Db 8124 GCGAGCGCGCACCC-CCGCGCTCCGCTTATCATGACACCCGAGCGGCTCCA-----CG 8078
QY 2167 oGlyAlaSerCysProValLeuAspLeuArgArgProPheAspAspLeuTyrLeuProP 2187
Db 8077 GGGTACGCTCGCTTCCAGCGGACGTACAGCGCGCGCGCTTACGAA----- 8032
QY 2187 oProAspHisGlyAlaProAlaArgGlySerProHisSerGlnGlyGlyLeuAspArgSer-- 2206

```
Db 8031 -----TCCGAGCGGCTCACACCACTCCAGAGCGCTCCGCGCACAT 7985
Qy 2207 -----ProGluProLeuThyThrsValLeuGlyGly 2218
Db 7984 CCCACTCGCACTCGGGGCCCACTCCATATCCCTCA-----GATGTC 7940
Qy 2218 YGluAaPrgIyIleGluProValSerProGluGlyMetThrGluProGlyYHisSerX 2238
Db 7939 GCGCACTGGTTTG-----CTCGGCGATTCAAGCCCTCGTAACCTCAC 7898
Qy 2238 gSerAlaVal---TyrProLeuLeuTyrAaPrgIyGluGlnThrGluProSerAArgMe 2257
Db 7897 TCTGCGCTCGCCACCTCCACTG-----CCACCGCGGCC 7865
Qy 2257 fclYserIyYSerProGlyAaThrsSerGlnProProAlaPhePheSer----- 2273
Db 7864 GGGGTCCTCTCCAGCTCGGCTTCAACACACTCCGCGAGGAC-GCCTCCCTCGGGTACTC 7806
Qy 2274 ---LysLeuThrGluSerAaSerAlaMetValYsSerIyYsGlnGlyIleAaIy 2292
Db 7805 CGCTGCGCTCCGCTCCACTCTCCACAGAGCGCGCTCTCTCCACGACATCAACA 7746
Qy 2292 slySleuAaThrHisAaAaPrgAaGlnProGluTyrAaPn----- 2305
Db 7745 CAGCTCAGACAAAGTCCGGCGCTCGAAGCACACAGTTCGCGAGAGCTCCGCGCTG 7686
Qy 2305 ----- 2305
Db 7685 CTCGACCATCCGCTCCGCGCGATCATCAAGCCCTGCGCTGCTAGTACGGCGCAG 7626
Qy 2306 ---IleSerGlnProGlyThrGluIlePheAaMetProAlaIleThrGlyThrsIy 2324
Db 7625 CTGCCATTTCAGCGCCCGGAGCGGCATCATC-----GTCAACGGAGT----- 7583
Qy 2324 uMetThrTyrAaSerGlnAlaValGlnGluHisAaSerThrsMetGlyLeuGluAl 2344
Db 7582 -----GAGTCACCTCGACGATGTACGGGTGCGGACACACAGCTCGTCCACAT 7533
Qy 2344 alIleIleArgIyAlaLeuMetGlyYsTyrAaPrgIyNTPrGluGluSerProLeuA 2364
Db 7532 CGTGGTGAAGCCCGCATCCCGGGGTAAGTTC-----TCGAAGACGACAGGCTCTC 7482
Qy 2364 rAlaAaAlaPheAaPrgLeuAaAlaSerAlaSerLeu-ProAlaIleMetProIleT 2384
Db 7481 GAACAAAGCG-----GGGCCA-----GGGGGAACCTGGCTCACGGCGACCTCCA----- 7435
Qy 2384 hTAlaAlaAaPrgIyAaSerAaPrgHisThrsLeuThrsProGlyGlyGlyAla 2404
Db 7434 -----CCAGGGGCGACGCTCATATGCTCGCTCGCTCTGGA-----GCCA 7395
Qy 2404 yValSerGlyAaPrgSerAaArg-----LysAlaYsSerProAlaProGlyLeuA 2422
Db 7394 GGTTCACAGC---CCTTCAGCCAGGAGGCCCGAGGAGCTCGACGGGCGAGCGCACTG 7338
Qy 2422 lAsSerIyAaPrg-----ProProSerValSerSerValHisSerGluGlyA 2438
Db 7337 CACGGGGATGCGCTTGATGAACAAGCCCAATCGCTCGATCCGGCAAGCTCCGGGG 7278
Qy 2438 sPcYsAaAaPrgAaThrProLeuThrsAaAaPrgValTPrGluAaPrgProSerA 2458
Db 7277 CCGCGCGAGACGCTCAACCCGAACAGAGAGT-----CGTCCCTCCGCTATA 7230
Qy 2458 lYserThrProPheProTyrAaPnProLeuIleMetAArgLeu-----GlnAla 2474
Db 7229 GCGCCACAGACAGCGCGCCAGCGCTTGCAACAGCGTCCCTCCGTACGCCCTGCTG 7170
Qy 2474 lYValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyP 2494
Db 7169 GCGCGAAGAGCGCTTCAGAGCGCGCTCTCTCGGCTCATCGGAGACAGCACTTCACC 7110
Qy 2494 rO-----HisHisAla 2497
```

```
Db 7109 CAGAGCGTCATACATACGCA 7090
RESULT 7
US-09-540-9597/c
; Sequence 9597: Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 9597
; LENGTH: 14462
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-540-9597
Alignment Scores:
Pred. No.: 7,98e-23 Length: 14462
Score: 698.00 Matches: 652
Percent Similarity: 32.50% Conservative: 301
Best Local Similarity: 22.24% Mismatches: 1028
Query Match: 5.28% Indels: 964
DB: Gaps: 143
US-09-522-753-5 (1-2517) x US-09-902-540-9597 (1-14462)
Qy 44 HighIsSerAaPrgTyrAlaSerHisLeuSerProGlySerIleIleGlnProGlnAa 63
Db 14437 CACCCAGCGCTCCAGCGGATGCTCGGCTCAGCGGACATCAAGCCCGCTTC 14378
Qy 64 ArgAaPrgSerLeuLeuSerGlnPheGlnProGlyAaGlnAaPrgSer-GlnGluLeuHis 83
Db 14377 AGAATCCCAAGCGCGCTCAACCCAC-----TCAACAGACGCTCCGCGCAATGCC 14324
Qy 83 sLeuAaPrgProGluSerHisSerTyrLeuProGluLeuGlyYsSerGluMetGluPheI 103
Db 14323 ACTCGACCTCGAGG----- 14308
Qy 103 eGluSerIyAaPrgProAaGluLeuLeuProAaPrgProLeuAaPrgProSerPro 123
Db 14307 -----CCACTCCATT 14297
Qy 123 uLeuAlaThrGlyGlnProAlaGlySerGluAaPleuThryAaPrgSerLeuThrsI 143
Db 14296 CCCCTCAGGTGATGCGCCCAAGCTGTTGCTCGGCGATTCAAGCGCTGTAACCTCAC 14237
Qy 143 ylySleuGlnProValSerProProSerProProHisThrsAaPrgProGluLeuGly 163
Db 14236 TCCCTCGCCCACTCCACTGCCACCGCGCGGAGTCTCTCCACTCGCTTCAACAGC 14177
Qy 163 lProPro-AaArgLeuSerIySglGluLeuIleGlnAaMetAaPrgValAaPrgGlu- 182
Db 14176 TCCGCGAGGGA-----CGCTCCCTCGGGTACTCCGCTCCG 14141
Qy 183 -----IleThrsValGlnGlnIleSerIySleuYslySlyGlnGlnIleuG 201
Db 14140 CTCGCCCTCCACTCTCCACAGAGCGCGCTCTCCACCGACATCAACAGCACTCG 14081
Qy 201 lGlnGluAlaAlaYsProProGluPrgGluYsProValSerPro----- 216
Db 14080 CCCAGGC-GCGTCTCGGATTCGCGACCGCGCTTCCACAGACAGCACTGCTCAG 14022
Qy 217 -----ProProIleGluSerIyHisAaPrgSerLeuValGlnIleIleTyrAaPrgIuA 234
Db 14021 CAGCGCTCGCGCTTC-GCCTGACAGAGTGGCTGTGATTC----- 13976
```

QY	234	enArglybValaGluAlaAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluL	254
Db	13975	-----ACTGACGCGCGGACACC-TTCCGGACGCTCGACGAACTCGACAGTGAAC	13922
QY	254	eupProLeuTyTAsnGlnProSerAspPheArgGlnTyTnIsgluAemIleLeuG	274
Db	13921	TTCCGACGCTGAATCCG-----CGGCTGCATC	13895
QY	274	lnAlaMetArgLybLybLeuIle---LeuTyTPhelybAArgAghnIbAlaTybG	293
Db	13894	ACCGTCAGCGCAATCCCTGTCCGACGTTCACTTCGGGACAGCGCGTTCTGAAGTC	13839
QY	293	ln-----TrrLybGlnbSerPheCybGlnArgTyTAsnArgIleuMetGlnAlaLeuGlu	311
Db	13834	AGGGAAGACTTGAACAAACAGGGCTCGACTAGGTCCGCTCGGCTGCAATGCCCTTCACC	13775
QY	311	yelybValaGluArgIleGlu-----AsnAsnProA	321
Db	13774	AGCTTCTCGAAGCGGAGCTCGTGGTGGCGTAGGCGCCAGGGTCACTTCACGCACTCG	13715
QY	321	rgArgArgAlaLybGluSerTybValArgIuTyTArgLybGlnPheProGlnIleA	341
Db	13714	GCCAACTATGCCGGAAGAAAGCTCGGACGTACGCGACCTTCAGACACAGAGTAATG	13655
QY	341	rgLybGlnArgTybLeuGlnGlnIuArgMet-----GlnSerArgValaGlyGlnArg	358
Db	13654	GCGAAATGGCCGATGAGCCCTCGCTCGGCGCTCGCGTCCGACCCCGATGGGAAGCC	13595
QY	358	LybSerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValaSerGluIleIleArg	378
Db	13594	ACCGTGATGTCGTCCTCCCGGAGAGCCGACCAACACCGCTCG-----	13551
QY	378	LyLeuSerGlnGlnGlnAlaSnLeuGlnIuArgIn---MetArgGlnLeuAlaV	394
Db	13550	--ATACGCCCGCCAGACAGACCAATGAACAGCGTGGCCCTTCTGCGCGCGGTCC	13493
QY	394	allieProPheMetLeuTyTAspAlaAspGlnIlnArgIleTyTPhelIleAsnMetAsn	414
Db	13492	ACTTGGCGGCA-----CAGCGCTTCGG-----	13470
QY	414	LyLeuMetAlaAspProMetLeuValTyTAspArgGlnValMetAsnMetTrrSer	434
Db	13470	-----	13470
QY	434	lunGlnIuLybGluThrPheArgGluTybPheMetGlnIbProLybAsnPheGlyLeuI	454
Db	13469	-----CCACAT-----	13464
QY	454	leAlaSerPheLeuGluArgLybThrValAlaGluCybValLeuTyTTyTLeuThrL	474
Db	13463	-----CTGC	13460
QY	474	yelybAsnGluAsnTyTLebSerLeuValaArgArgSerTyTAsnArgArgIlybSer	494
Db	13459	CGCGAATTCACCTCGAAGAGCTGTGCGCGCGGCGGAGAGCTGTCTGTGGAGGTCC	13400
QY	494	lnGlnGlnGlnGlnGlnGlnGln-----	502
Db	13399	AGGAATCGCGGCGGCGCGCCAGACACTCGCGCAGAGGAACTGCGATTCCAACACC	13340
QY	503	-----GlnGlnG	505
Db	13339	TGCTCTTCAACACTCGCGCTCGCAGAGCGGTAGTCCGCGTACGCAAGGAGCTCC	13280
QY	505	lnGlnGlnGlnGlnGlnPrometProArgSerSerGlnGlnIuLybAspArgIuLybGluL	525
Db	13279	GCGAAATGTACGCGCAACCTTG---CGAAGACGCGTCGAGAGCGCGCCCAATCCCGA	13223
QY	525	ybGluLybGluIuArgIuLybGluGluIuLybProGluValaGluAsnAspLybGluAspL	545
Db	13222	AGCGACACGCTCGCGACCAACCGTCCGAGACATGTATGTAACAGCAACAGAGAGG	13163

QY	545	euleuLysgluLysThrAlaAspHisSerGlyGluAspAsnArgLysgluAlaValA	565
Db	13162	TGGC-CTGCTGGGAGCTTCAACAGAGG-----	13133
QY	565	LaserysGlyArgLysThrAlaAsnSerGlnGlyArgLysArgS	585
Db	13131	-----GGGCGACGACGGCCCAACGGGAGCTGAAATGGCGGTCGACTTCCGGCG	13081
QY	585	erMetAlaSerGluAlaAsnSerGluAlaIleThrProGlnGlnSerAlaGluLeuA	605
Db	13079	CA-----CCCGGTCGGCAGC-TCGGACTCATCTGGGCGCCACACTCCACACGAGAGGT	13024
QY	605	LasMetGluLeuAsnGluSerSerArgThrThrGluGluGluMetGluThrAlaLys	625
Db	13023	GGACCGGGGAT--GGCGAAGCCATGGCTGTGACAGTGTCCGCTTCATTCGGGAG-	12966
QY	625	ysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaTrgMetValGlySerIet	645
Db	12967	-----GCCATGCCAGACTTCAATGACGGCGCA	12944
QY	645	hrValSerGlnCysLysAsnPheThrPheAsnTrpLysLysArgGlnAsnLeuAsp----	663
Db	12939	CCAGCTCCGAG-----AAGCTCGCTCCAAAGCAGAAACAT	12904
QY	664	-----GluIleLeuGlnGln-HisLysLeuLysMetLysgluArgAsnAlaArg	680
Db	12903	CCAGGGGGCCCTCCACCCCGCAGCACAATCCGTCGTGTAATCTGCATGACCCCGCGCTTCA	12844
QY	681	ArgLysLysLysLysAlaProAlaAlaAsnSerGluGluAlaA--PheProPval	699
Db	12843	GCCATTCAGAACACACATGCTGCTGCGCGAATGACAGCGGACCCGGTCCCGCCGAG	12786
QY	700	ValGluAspGluGluMetGluAlaAsnSerGlyValSerGlyAsnGluGluMetValGlu	719
Db	12783	GCATCGGACCGAGCGGCGCAGACGCTGCTGACTTC-----	12755
QY	720	GluAlaGluAlaLeuHisAlaSerGlyAsnGluVal-----ProArgGlyGlu	735
Db	12750	GAGTCGCTGCTCCACTGCTGGGGCAAGACCGCAAGTGGCGGCTCGAAGAGGGCCG	12691
QY	736	CysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerProHis	755
Db	12690	GTAACGGGACCTTACGC-----CCAGACCGGCAC	12661
QY	756	ThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysAspProAlaThrIleGluAlaA	775
Db	12660	GCAACAGCG-----CCACACCTCGCGGCACATGAGCAT	12625
QY	776	AspGlyPro-----ProProGlyProProThrProProArgArgThrSer	790
Db	12624	GACCAACCCGACGCGAAGATGTCGTCTCCGGGCCACCGCTCCACCCCAAGA-----	12571
QY	791	ArgAlaProIleGlu-----ProThrProAlaSerGlu	801
Db	12570	CTGGGAGCCAGACAGCGGCGCAACGACTTTCATGGGGAGAAAGCGCGCTCATAGCGG	12511
QY	802	AlaThrGly--AlaProThr-----ProPro-----	809
Db	12510	GAGCCGGGACACGCGCGGACAGCTTGGCAGCGGCGACGCGCGCCACAGTCCACTTGGCAT	12451
QY	810	-----ProAlaProProSerProSerAlaPro-----	818
Db	12450	TGGGCGTCAAGCGGAGGGCGTTCAGACACACACGCGCCACGACCAATGTGCCAGAA	12391
QY	818	-----	818
Db	12390	GCCGCTGTGAGAGAAACCCCTGAGGGCGCGATGTCGGAGCCGCGCTGGCCGTACGAA	12331
QY	819	-----ProProVal-----ValProLys	824
Db	12330	CGGCGTACCCACACGCGCATTTGCTCCAGGACACTCTTCTGCGCGGACACACACCGCT	12271
QY	825	GluGluLysgluGluGluThrAlaAlaAlaProProValGluGluGlyGluGluLys	844

```
Db 12270 CCTGACGCGGGAGCGCTGTGAGACCGCGCACTTCT-CCGGGGTCCACCGGAAAC 12212
      :::
Qy 845 ProProAlaAlaGluGluLeuAlaValAspThrGlyValAlaGluLeuProValIysSer 864
      |||
Db 12211 CCT-----CTCACCTTCACTGGAAAGTCCGCGCGGATGAAGTCC 12170
      |||
Qy 865 Glu-----CysThrGluGluValaGluGluGlyProAlaIleYsGlyAspAlaGluAla 882
      |||
Db 12169 AGTGGCCATCCCGACACCGCACCGCACTTGTCCCGTGGGTACATCTCTGCTCCAGCC 12110
      |||
Qy 883 AlaGluAlaThrAlaGluGlyAlaLeuYsAlaGluYsLeuGluGlyIysSerGlyArg 902
      |||
Db 12109 TCCATGCTGACGAGCGGCGGACGTACGCTCGCGCGTC-----AACTCCGGGCGC 12059
      |||
Qy 903 AlaThrThrAlaIysSerSerGlyAlaProGluAspSerAspSerSerAlaThrCysSer 922
      |||
Db 12058 CATACGTACCCCTCGGTGACCTGGCCCTCCACGAAACACTCTCCCGCCACGCGCCAC 11999
      |||
Qy 923 AlaAspGluValAspGluAlaGluGlyIAspYsAsnArgLeuLeuSer-ProArgPr 942
      |||
Db 11998 -----GGCACGAGACGAGGCTCCGGTCCAGACGTAA 11966
      |||
Qy 942 oSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGluIys-----Pr 960
      |||
Db 11965 ACCCGGTGTGCGCATCGGCGGTCCCA-----GGGCAACCGTCCGCGGTGACTCC 11915
      |||
Qy 960 oLeuAspLeuYsGlnLeuYsGlnArgAlaAlaAlaIleProProIleGlnVal----- 978
      |||
Db 11914 GCTGACGCTG-----GCAGCTTCACCTCGCCCGCCGACGACACCCACCGCTC 11870
      |||
Qy 979 -----ThrYsValHisGluProProArgGl 987
      |||
Db 11869 GTCTCCGTGGGGCCGAGTGTGTGACCTGCGACCGCGGCGCCAGCCCGCGCACTGC 11810
      |||
Qy 987 uAspAlaAlaProThrYsProAlaProProAlaProPro----- 1001
      |||
Db 11809 TCCAGACGCTCCACGAGGCGG---CCTCTCTCCACACGACGCGCTTCCGTGGCAAC 11753
      |||
Qy 1002 -----ProGlnAsnLeuGln-----ProGluSe 1009
      |||
Db 11753 ACCTGCGCGGCGCGCGCCACTCATGACGCGCGGAGGTGCGGACGACCACTTCAAA 11693
      |||
Qy 1009 rAspAlaProGlnGlnProGlySerSerProArgGlyIysSerArgSerProAlaProPr 1029
      |||
Db 11692 CACTCCACCCCATCGCGCTGGAAGTACTCTCCACTCCGCGCGGTGCTGCGCTCTCC 11633
      |||
Qy 1029 oAlaAspYsGluAlaAlaAlaGluAlaGlnYsLeuProGlyAspProProCysTr 1049
      |||
Db 11632 TCGGTACAGACATGCA---GCAGTCCCCCGGTCCACAGCGCGCGTAACAGCACCGGT 11576
      |||
Qy 1049 pThrSerGlyLeuProPheProValProProArgGluValIleYsAlaSerProHisAl 1069
      |||
Db 11575 C-----CCAGGTCCCGCGCCACCGTCCGACACGAGCGGAAAGCTGTGACCTCC 11528
      |||
Qy 1069 aProAspProSerAlaPheSerTyzAlaProProGly----- 1081
      |||
Db 11527 GCCAACCCCAAGCGCGCATTCGCGGCTTCCAGCTAATGGGCAAGCTCCGTTGTGACC 11468
      |||
Qy 1081 ----- 1081
      |||
Db 11467 GCCAGCGCTTGGGCGCGCGCGTTCGCGATGTGAACAGACGATACCAGTTGCTCG 11408
      |||
Qy 1082 -----HisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArg----- 1098
      |||
Db 11407 GGTTCGACCTTAACCGCGAGTGTCTCGTGGCTGGCTTCGAGCGCGCGGTGTCTCA 11348
      |||
Qy 1099 -----ProPro-----ThrIleSe 1103
      |||
Db 11347 TCCATCTCAACCGCGCGCGGACGACTCCCTCGAAAGTCCCGGCATGCCGCGCTGCGT 11288
      |||
Qy 1103 rAsnProPro-----ProLeuIleSerSerAlaYsHisProSerValleuGlnArgGl 1121
      |||
```

QY 1403 lYpRoLeuLybLeuLybProAlaHISgluGluLeuValAlaThrValGluAlaGly 1423
 Db 10278 -----GCACCCATGACACCGCGCT-----GGACACAGCTCCCGCGCTGGG 10239
 QY 1423 rGserIleHISgluIleProArgGluGluLeuArgHISThrProGluLeuProLeu- 1441
 Db 10238 CTGTGCGGTGGGTAGCTCCAGGTGAGGAGCGCCGACAGACGTGCTGGCCGACAGTC 10179
 QY 1442 --AlaProArgProLeuLybGluGlySer---IleThrGlnGlyThrProLeuLybGly 1460
 Db 10178 GACCTGCGCTCTCCAGACACGCGCGCTCCAGCCATTGCGCTGCCACCGCTAGTCCGC 10119
 QY 1460 sPThrGlyAlaSerThrThrGlySerLybHISAspValArgSerLeuIleGlySerP 1480
 Db 10118 GTATGACAGCGGACGCTCCGCGACACCGGAGACCGCTTCCGA-----CGAAGGC 10068
 QY 1480 rGglYArgThrPheProProValHISProLeuAspValMetAlaAspAlaArgAlaLeuG 1500
 Db 10067 CGCATAGAGCGCGCCGACCTCGCGACCA-----GCACGCGCCAGCGACGACCATC 10017
 QY 1500 LuArgAlaCybTyArgGluGlySerLeu-----LySerArgProGlyThrAlaSerSers 1518
 Db 10016 CGAGACGATGTGATGACAGACACAGACAGCGGCGTGACGTCCGCGACAGCGCGACAG 9957
 QY 1518 ex-----GlyGly-----SerI 1522
 Db 9956 CCCACGCGGACACAGGCGGCGCGCTCCAGATGAACGCGCGACGCGCTTGCAGTCCG 9897
 QY 1522 LeAlaArgGlyAlaProValIleValPro-----G 1532
 Db 9896 ACCGCGCGGGAATTGGCTCGCGCTCCATTCCGCGACGCTGTCAAGTCTCGAAGGC 9837
 QY 1532 LuLeuGlyLybPro-----ArgGln 1539
 Db 9836 GACCTGACCTCATGACGCGGATGATGATGACCTGACAGCGCTGGCTGTGGGAGGACAA 9777
 QY 1539 exProLeuThrTyArgLysAspHISglYAlaProPheAlaGluHISLeuProArgGlySerP 1559
 Db 9776 CGTCGCGCGCA---ACGACTCATGGCGCGCACCC-ACGCGGAGAACCGCGCTGCACAG 9721
 QY 1559 roValThrMetArgLysProThrProArgLeuGlnGlySerLeuSerSerSerLybAla 1579
 Db 9720 CGCTCACATGAGAGAGCCCTCGATCCGAG-GACGATGGGGAAGTTGTAGCTGGCACTG 9662
 QY 1579 lAserGlnAspArgLybLeuThrSerThrProArgGluIle-----AlaL 1594
 Db 9661 CCCGATGGAG-----CTGTGATGATGAACAGAGCCGCTGTGGCGAAGACAGACGGGC 9608
 QY 1594 yAserProHISserThrValProGluHISHisProHIS----- 1606
 Db 9607 AATGCGCCCATGGAGGTACCCGGGTGACGCGTGGACGCTCGGCGCGCGCTTCCCGC 9548
 QY 1607 -----ProIleSerProTyArgLys 1613
 Db 9547 GCCGACCTCTCGATTGCGACGCGGAGCGCTTCCAGACGACAGTGAACGGAAGAGGTCCG 9488
 QY 1614 -----LeuLeuArgGlyValSerGlyValAspLeuTy-----ArgSerH 1627
 Db 9487 AGGAGGACATCCACGCGGAAAGCGAGCGGATGCGCGGATGAGCTGCGCGACAGAC 9428
 QY 1627 lAsIleProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAla 1647
 Db 9427 GAGTGCCTCCGCGCACTCGAAGATCAGGCTC-CGCGCTCACCGCGCTGATGCC----- 9375
 QY 1647 lAlaIleTyThrLeuProArgHISLeuAlaProAspProThrTyProHISLeuTyProp 1667
 Db 9374 -----CAGCACTTCACACCA----- 9360
 QY 1667 roTyThrLeuIleArgGlyTyProAspThrAlaAlaLeuGluAspArgGlnThrIleAla 1687
 Db 9359 -----GATGATGCTCAGACTTCTGCTCG 9338

QY 1687 snAspTyThrIleThrSerGlnGluMetHISHisAsnThrAlaThrAlaMetAlaGlnArg 1707
 Db 9337 GTGAC-----CGTACGCGGC 9223
 QY 1707 lAspMetLeuArgGlyLeuSer---ProArg-----GluSerSerLeuAlaLeu 1723
 Db 9322 TCGATGTGCTGACGGGGCTCGGAAGCGCCGCTCGGAGGCGCAACATGCTTGGCGTGC 9263
 QY 1723 snTyAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHISLeuProVal 1743
 Db 9262 ACCGTGCGGTGGAGCTGAC-----CGGACGCGCTTCACACCA-CCA----- 9220
 QY 1743 euValProProThr---ProGlyThrProAlaThrAlaMetAspArgLeuAlaTyLeuP 1762
 Db 9219 -----CCACGAGCGGACCATGTACTCGGCACTTCTGCTCAGT-----ACTC 9174
 QY 1762 roThrAlaProGlnProPheSerSerArgHISerSerSerProLeuSerProGlyLyP 1782
 Db 9173 CTTCAGGCGCCCTCCCTCTCTCCGCGCGCACAGTACGCTT---CCACCGCTTCTC 9117
 QY 1782 roThrHISLeuThrLybProThrThrThrSerSerSerGluArgGluArgAspArgAsp 1802
 Db 9116 CTGCGCTCTCTCCCTACACCGCTGCG----- 9088
 QY 1802 rGglYArgAspArgAspArgGluArgGlyLybSerIleLeuThrSerThrThrValG 1822
 Db 9087 -----CTTCCTTACCTCGCGCTGCTG 9066
 QY 1822 LuHISAlaProIleTrpArgProGlyTyThrGlnInsSerGlySerSerG 1842
 Db 9065 CTTCAGGCGCTCTCCACCTCCGCACTCA-----TCGGTACCCCTCACTTCCAC 9012
 QY 1842 LyGlyGlyGlyGlySerSerSerArgProAlaSerHISAsnHISAlaHISGlnHISerP 1862
 Db 9011 CTGCTGTCCTCCAGCGCCCAAGAACTCCAGACCCCACTCTCTGCACTTCACTCTGTC 8952
 QY 1862 roIleSer-----ProArgThrGlnAspAlaLeuGlnGlnArgProS 1876
 Db 8951 CCCCGTCGGTACACCGGCTCCCGCTCCCAACGGTGAAGGACGAAAGCTCCGCGT 8892
 QY 1876 exValLeuHISAsnThrGlyMetLybGlyIleIleThrAlaValGluProSerLybProT 1896
 Db 8891 CCACTCCGCGCTCC-----CCACGTACCCCTCCGCGACCCCTC 8853
 QY 1896 hrValLeuArgSerThrSerThrSerSerPro-----ValArgPro 1910
 Db 8852 TCTCTCCAGTACACTTCTCCCTCAACCCGACCGGACCGAGCATTCCTTGTCAG 8793
 QY 1910 lAlaIleThr-----PheProAlaIleThrISys-----ProLeuG 1922
 Db 8792 CAGCTACACCGCGGTATTCTCCACCGGCTTCCAAATGGGACCGTCAACCCCACT--- 8737
 QY 1922 LyGlyThrLeuAspGlyValTyProThrLeuMetGluProValLeuLeuProLybGlu 1942
 Db 8736 -----CTCTCTGCGAGCGCGGTGACGTGCTGAAAGTCTG 8700
 QY 1942 lAsProArgValAlaArg-ProGluArgProArgAlaAspThrLybHISAlaPheLeuAla 1961
 Db 8699 CCCCTGTGCGCGGTACCCGTTGAGAGATGCCCCCACTCTCACTT---CTCTCT 8643
 QY 1962 LyAsProAlaArgSerGlyLeuGluProAlaSerSerProSerLybGlySerGluPro 1981
 Db 8642 CACCCGCGCCACGACAGACAGCTCTCTC-CGCGGACACTTGCCTCAGCTCACTAGCG 8584
 QY 1982 ArgProLeuValProProValSerGlyHISAlaIleThrIle---AlaArgThrProAlaLyb 2000
 Db 8583 CTTCGCGGTGTAGCTGCTGATCTGCTGAAAGAGCTGCGCTGACAGAGAGCTCA 8524
 QY 2001 AsnLeuAlaProHISHisAlaSerProAspProAlaProProAla---SerAlaSer 2019
 Db 8523 CCCCTGCTCTTCAAGCTTCCGCGCGGCTTCCAACTCCACAGGCGCTTCCGGGTACA 8464
 QY 2020 AspProHISArgGlyLybThrGlnInsLybProPheSerIleGlnGluLeuGluLeuArg 2039

[illegible][illegible]

QY	1161	uAlaIaProPheSerGlyValIleGlnGlnGluGlnLeuSerProAlaGlyGlnAlaGlyProPhe	1181
Db	839496	TECCGCCGTTGCCCGCGTTCGCGATCAACACCGCGGTCCGCCGCCGCCGCCGCCGCCAC	839437
QY	1181	oGluSerLeuGlyValProThrAla---GlnGlnAlaSerValLeuAlaGlyThrAlaLe	1200
Db	839436	C-----GTCGCCGCCGCCGTCTCGCGCGCGCTTGGCCGCCGCTTGC	839389
QY	1200	uGlySerValProGlyGlySerIleThrIlys-----GlyIleProSerThrArgVa	1217
Db	839388	TECCCATCAGCCCGGGGGGCCGCCGCCGCCGCCGCCGCCGTCTGCCGACCTTTTAC	839329
QY	1217	lProSerApsSerAlaIleThrTyThrGlyGlySerIleThrIhIsglyThrProAlaApsVa	1237
Db	839328	TAATCCCGCGCGCGCGCGCGCGAGCCGCCGAGCGGAAGAACATGACCGCGTCC--	839274
QY	1237	lLeuTyrlsGlyThrIleThrArgIleIleGlyGlnApsSerProSerArgLeuApsAr	1257
Db	839273	-----GCCGCTCCGCCGCCGCCGCCGGAATCCGCCCG	839242
QY	1257	gGlyArgGlnApsSerLeuProLysGlyIhIValIleTyrgLuglyLys-----	1273
Db	839241	TECCGACTAGCGCGCGCGTCCGCCGCCGCCGCCCATCAACAGTGCAGATGTCCGCCGT	839182
QY	1274	---LysGlyIhIValLeuSerTyrgLuglyGlyMetSerValThhGlnCysSerIysG	1292
Db	839181	TECCGCCGCCGCCGCCGCCG--CCCGCCAGTGAAGTTCGCCGACCCGCCGTGCCGCTTC	839123
QY	1292	uApsGlyArgSerSerSerGlyProProIhIsgluthrAlaAlaProLysArgThrTyra	1312
Db	839122	CTTCCGCCGCCCAACACGCGCGCGCGCGCGCGCGCGCGCTGCCCGCGTGGAC--	839055
QY	1312	pMetMetGluGlyArgValGlyIArgAlaIleSerSerAlaSerIleGluGlyLeuMetG	1332
Db	839064	----GCGCGTTCGAGATCCGCGCAGCGCGCGACCCGCCGCCGCCGGAACATGCGC	839009
QY	1332	yArgAlaIleProProGluAlaArgIhIserProIhIshIleuLysGlnGlnIhIshIleAr	1352
Db	839008	GGGTTTCCGCCGCCGCCGCCG--CCCACTCCGCCGCCGCCGCCGTCGTGGAG	838955
QY	1352	gGlySerIleThrGlnGlyIleProArgSerTyValGlnAlaGlnGluApsTyLeuAr	1372
Db	838954	G-----CCCGCGCGCGCGCGCGCTGCCGGAACGCCCGCGCC	838916
QY	1372	sArgGlnAlaLysLeuLeuLysArgGlnGlyThrProProProProProSerArgAs	1392
Db	838915	CCGCCGTTACCGCGCGCGCGCGCGCG--GGGGGCCAAGCCGCCGCCGCCG--	838865
QY	1392	pLeuThrGlnAlaTyTyLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaIhIsg	1412
Db	838864	-----CCGTTG-----	838859
QY	1412	uGlyLeuValAlaIhThrValIleGlnGluAlaArgSerIleIhIsglIleProArgGlu	1432
Db	838858	-----CCGATCAACACCCGCCGCCGCCGCC	838833
QY	1432	uLeuArgIhIshThrProGluLeuProLeuAlaProArgProLeuGlyGlySerIleTh	1452
Db	838832	GTTGGCTCCGCGTCCGCCGCCGCCGCGCGCTTGGCGCGCTTGC	838790
QY	1452	rgInglTyhProLeu-----LysTyArgAspTh	1461
Db	838789	-AGCGGCGCCCGCGTGGCCGACGAATTCGCGTTGATCGGGGCGACGACGCGCGAC	838731
QY	1461	rgLyLaseThrThrThhGlySer---LysLysIhIshArgValArgSerLeuIleGlySerPro	1481
Db	838730	GGCGGCGGCTCGCGCGCGCGCAACGCGCTCCGCCGCGCTCAGGACCTGACGAATG	838671
QY	1481	lYArgThrPheProProValIhIshProLeuApsValMetAlaSerAlaArgAlaLeuGlu	1501
Db	838670	GGCGTAAACGCGCGCG--CTCTG--	838648
QY	1501	rgAlaCysTyrgLuglnGlnSerLeuLysSerArgProGly-----ThrAlaSer	1517

Db	838647	-----CGCTGGCGGCTGGTAAAGCTGGCCCTTGGGACCAACGAGCGCCG					8386
Qy	1517	erSerGlyGlySerIleAlaArgLysAlaProValIleValProGluLeuGlyLysProA					1537
Db	838601	AATGGCTGTGCACACCTCATCTGCGACCGCGGGGTGCACACATCCGTGATTGACGGCG-					8385
Qy	1537	rgGlnSerProLeuThrTyGluLysAspIleGlyAlaProPheAlaGlyHisLeuPro---					1555
Db	838542	-----CCGCGCCATTAGCCGAGCT					8385
Qy	1556	-----ArgGlySerProValThiMetArgLysLysProThrProArgLeuGlnGluG					1572
Db	838573	AATGGCGGACCCGATCCCGGCAATCCGCTGGCGCGGACCAACGAGCTCCGGCGAAAC					8384
Qy	1572	LySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSer-----					1587
Db	838463	CATCACCCATGACATCCGATCCCTCCCA-----CCCTCAGCGCTGGGCCGACTA					8384
Qy	1588	-----ThiProArgGluIleAla-----LysS					1595
Db	838409	CCAGCGACACACCCGTCGGGTGGATTAATTCGAGAAATCAAGATATCGCATCCG					8383
Qy	1595	erProHisSerThrValProGluHisAspHisProIleSerProTyGluHisLeuI					1615
Db	838349	CACCGCATGACGCGCTTCACACGTTCCGAG-----GCCGACGCTGTGACGCG					8382
Qy	1615	euArgGlyValSerGlyValAspLeuTyTrArgSerHisIlePro-----LeuAlaP					1632
Db	838288	TCAGGGGTGGGGGGGGGCTTCGCAACAAAGGGGGGGGGTCCGGGACAGAGGGCGCGG					8382
Qy	1632	heAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyTrLeuP					1652
Db	838238	CGAAACCCGCA-----CCGCGCGGATAGCGCTGTGCACGACGACCCGCG-----					8381
Qy	1652	roArgHisLeuAlaProAspProThrTyTrProHisLeuTyTr-PropoTyTrLeuIleArg					1671
Db	838194	--ACGGGCTCTTGACAAAGCGCTGACAGCGCGGGGATCTTCGCGAATATCTGGCGCG					8381
Qy	1672	GlyTyTrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyTrIleThr					1691
Db	838136	CG-GCCCCATTGGGGCGCGCG-----TCC					8381
Qy	1692	SerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArg					1711
Db	838110	GCAAGCCATCTGCATCACCGCGGAGTGTGACCGCCGGAAGCCATCCGTGCCCGCGG					8380
Qy	1712	GlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyTrAlaAlaGly-Pro-----Ar					1729
Db	838050	CTACGGCAACCCGTTGAGCCCGTTTGGCCACGACGAAACCCGCGCGCGCGGCTTCC					8379
Qy	1729	gGlyIleIleAspLeuSerGlnValProHisLeuProValIleuValProProThrProgl					1749
Db	837990	GGGGCTGGCGGCGTGGCGCGCGGCGCGCGGCTTCCGCGCGT---CCGCGGAGCGCAT					8379
Qy	1749	YThrProAlaThrAlaMetAspArgLeuAlaTyTrLeuProThrAlaProGlnProPheSe					1769
Db	837933	CAGCGCGGCGTGGCGG-----CCGACGCGCACCGGCGCGG-----					8379
Qy	1769	rSerArgHisSerSerSerProLeuSerProGlyGlyProThr-His-----LeuT					1786
Db	837899	-CCGAAGCCACGACGCGGAGCATTTGCGCGGCGCCACCGACCCACCGCGCGTCCGAA					8378
Qy	1786	hIlyProThr-----ThiThrSerSerSerGluArgGluArgAspA					1800
Db	837840	CCCCCGCAACCCCGCGCGCGCGCGCTGAGACACGACGACCCCGCGCTTCCGACCGGCCCC					8377
Qy	1800	rgAspArgGluArgAsp-----ArgAspArgGluArgGluLysSerIleLeuThrs					1817
Db	837780	GCGGCCCCACCGAGAAACCAAAATTCGCGCGCGCGCGCGCGCTGCGCAACAG					8377
Qy	1817	erThrThrThrValGlu-----HisAlaProIleThrPArgProGlyThrGluGln-Ser					1834

Db 837720 CAGCCCTGCGCTACCGCCGCGCCGCGTCCACCGGTGTCAGGTGTTTGGCCGTACCC 837661
 Qy 1835 SerGlySerSerGlySerSer-----GlyGlyGlyGlyGlySerSerSerArgProAla 1852
 Db 837660 GCGCGCTCCGCGCGCGCGCGCGCGCGCGAGATTCGCGATTCGCGCGCGCGCGCGCGCGCG 837601
 Qy 1853 SerHisSerHisValHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGln 1872
 Db 837600 AACCCCGCCGTCGATTAAGTCGAGCCCGCGGTCCGCGCGTGCAGCTCCGCGCGAGCC 837541
 Qy 1873 GlnArgProSerValLeuHisAsnThrGlyMetGlyGlyIleIleThrAlaValGlnPro 1892
 Db 837540 GATGACCGCGGCGC---CAGCAGTCCGCGGTCCGCGCGCGCGC---CCC 837502
 Qy 1893 SerIleProThrValLeuArgSerThrSerThrSerSer--ProValArgProAlaIleThr 1912
 Db 837501 ACCGCGCCCGCGCGC-----GCCACCAACAAGCGCGGTTCACGCGCGCGCGCGCG 837454
 Qy 1912 rPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValIleProThrIle 1932
 Db 837453 GCGCCCAACGCGCGCGCGCGCGCGCA-----CCGACCCC 837421
 Qy 1932 uMetGlyProValLeuLeuProGlyValAlaProArgValAlaArgProGluArgProArg 1952
 Db 837420 GCCTCCCGCGCGCGCGCGCGCGAGT-----ACGATGCTGCTCCGCGCGATCCGCGC 837370
 Qy 1952 gAlaAspThrGlyHisAlaPheLeuAlaValProProAlaArgSerGlyLeuGluProAl 1972
 Db 837369 GCG 837319
 Qy 1972 AsSerSerProSerIleGlySerGluProArgProAlaValProProValSerGlyHisAl 1992
 Db 837318 CCG 837262
 Qy 1992 AsThrIleAlaArgThrProAlaValAsnLeuAlaProHisHisAlaSerProArgProArg 2012
 Db 837261 CCGCG---GCGACCG 837205
 Qy 2012 oAlaProProAlaSerAlaSerAlaSerArgProHisValArgGlyLeuThrGlnSerIleProPhe 2032
 Db 837204 GCG 837177
 Qy 2032 rIleGlnGlyLeuGlyLeuArgSerLeuGlyGlyHisGlySerSerIleProGluGly 2052
 Db 837177 ----- 837177
 Qy 2052 yValGluProValSerProValSerSerProSerLeuThrHisArgIleValLeuProGly 2072
 Db 837176 -----CCG 837132
 Qy 2072 sHisValLeu-GluGlyLeuAspIleSerHisValLeuGluGlyGlyLeuArgProGluArgProG 2092
 Db 837131 -----TTGCGCATACG 837076
 Qy 2092 yProValIleValLeuGlyGlyGlyAlaValHisValLeuProHisValLeuArgProLeuProGlu 2112
 Db 837075 CCGCTGT---GTTTCAGCGGTGCGAGCGGAGAG-GCGCTGCGCGCGCGGTTCAGCGCGCG 837020
 Qy 2112 eGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValIleValHisGlnArgV 2132
 Db 837019 CCG 836982
 Qy 2132 aValIleThrLeuAlaGlnHisIleSerGluValIleThrGlnAspArgIleThrHisIleIle 2152
 Db 836981 -----GATACCG 836954
 Qy 2152 rGlnGlnIleuSerAlaProLeuProAlaProLeuIleArgPheProGlyAlaSerCysP 2172
 Db 836953 CGATGCGCGCGGTGCGCGCGCGGTGCGATGACACCGCGGTGCGCGCGCGCGCGCGCGCGTAC 836894
 Qy 2172 rValLeuAspIleValArgArgProPro-----SerAspI 2183
 Db 836893 CCGCGGTACAGCGCGCTTGCACCGCGCGCGCGCGCGCGCGCGCGTACGATCAGCAACC 836834

Qy 2183 eUTyIleProProProAspHisGlyAlaProAlaArgGlySerProHisSerGlyLeuGly 2203
 Db 836833 CCGCGCTTGCAGCG 836774
 Qy 2203 yIleValArgSerProGluProAsnIleThrSerVal-----LeuGlyGlyGlyGlu 2220
 Db 836773 CCGCGACCCACCGCGAGCCCAACACAGCC-GCGGTACACCGCGCGCGCGCGCGCGCGCG 836715
 Qy 2220 ePrgIle-----GluProValSerProProGluGlyMetThrGlnIleProGlyHisSerArg 2239
 Db 836714 TCGGTATCGCGCAACCGCGCGGTCTCCCGCGATCCGCGCGCGCGCGCGCGCGCGCGCGCG 836655
 Qy 2239 eAlaValIleProLeuLeuIleThrArgAspGlyGluGlnThrGluProSer----- 2255
 Db 836654 TTGCG 836619
 Qy 2256 -----ArgMetGlySerIleSerSerProGlyAsnThrSerGlnProProAlaPheHis 2273
 Db 836618 AGCGCGCGCTCCGCGGATCCACCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 836559
 Qy 2273 eTyIleValThrGluSerHisAsnSerAlaMetValIleSerIleValGlnGlnIleAsnIle 2293
 Db 836558 CAGCGCCCG 836529
 Qy 2293 yLeuAsnThrHisAsnArgAsnGluProGluIleThrAsnIleSerGlnProGly----- 2310
 Db 836528 AAGCGAGTCCG 836469
 Qy 2311 --ThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyArgSerG 2330
 Db 836468 CCAACGCGCTCGCGCGATCTCCCGCGAGCCCGCGGTGACGCGCG-CTAGCGCGCGCGCGCG 836410
 Qy 2330 ImlaValGlnGlnHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgIleValAl 2350
 Db 836409 GCGGTGCG 836393
 Qy 2350 eUmetGlyTyIleArgIleIleProGluIleSerProProLeuSer-AlaAsnAlaPheAsn 2369
 Db 836392 -----CCAAAGCG 836368
 Qy 2370 ProLeuAsnAlaSerAlaSerLeuPro-----Ala 2379
 Db 836367 CCG 836308
 Qy 2380 AlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrIleu----- 2394
 Db 836307 CAACAACAGCCCACTGATCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 836248
 Qy 2395 -----ThrSerProGlyGly-----GlyGlyIleValIleValSer 2406
 Db 836247 CCG 836188
 Qy 2407 -GlyArgProSerSerArgIleValIleValSerProAlaProGlyLeuAlaSerGlyAspArg 2426
 Db 836187 GCGCGAGCGCGAGCGAGCG 836131
 Qy 2426 gProProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThr 2446
 Db 836130 CCG 836086
 Qy 2446 rAsnArgValIleProGluAspArgProSerSerAlaGlySerThrProPhe---ProTyArg 2465
 Db 836085 GCGGTGCGCGATACCG 836026
 Qy 2465 nProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeuPro 2485
 Db 836025 CAGGTCTGTGTCACCTGCGCGCGGTGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCG 835967
 Qy 2485 oAlaGlySerGlyPro 2490
 Db 835966 CCGTGGCGTGAAGCG 835951

```
RESULT 9
US-09-103-840A-2/c
/ Sequence 2, Application US/09103840A
/ Patent No. 6294328
/ GENERAL INFORMATION:
/ APPLICANT: FLEISCHMAN, Robert D.
/ APPLICANT: WHITE, Owen R.
/ APPLICANT: FRASER, Claire M.
/ APPLICANT: VENTER, John C.
/ TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
/ TITLE OF INVENTION: TUBERCULOSIS
/ FILE REFERENCE: 24366-20007.00
/ CURRENT APPLICATION NUMBER: US/09/103,840A
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: Patentin Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 4403765
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ FEATURE:
/ OTHER INFORMATION: CDC 1551
/ OTHER INFORMATION: "n" bases at various positions throughout the sequence
/ OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 1,54e-19 Length: 4403765
Score: 685.00 Matches: 530
Percent Similarity: 32.37% Conservative: 186
Best Local Similarity: 23.96% Mismatches: 852
Query Match: 5.18% Indels: 652
DB: Gaps: 95

US-09-522-753-5 (1-2517) x US-09-103-840A-2 (1-4403765)

QY 546 LeuLysGluLysThrAspSerThrSerGlyGluAspAsnAspGluLysGluAlaValAla 565
DB 843605 GTTCGTCGGAAGGTGATGGGCGCTGGCGGTGACGCGCTTCGACGAAATGCCATCTCGC 843546
QY 566 SerLysGluArgLysThrAlaAsnSer 578
DB 843545 TCGCGGTGACGAATCTCGAAGATGCGTGATTCGTTGCCAGCGCGGAGAAACGACCCCA 843486
QY 579 579
DB 843485 TACACAGCTTCGGACATCTGAAGGCGTCGTCACGCGGAGACATCGATCGAACGAGGCGG 843426
QY 593 Glu 600
DB 843425 GACGGTTCGGGTGATTCGGGTGTCACCGGCGTGTGACGACGACGACGATTCGCGAG 843366
QY 601 601
DB 843365 AAGCAATTCGCGCTTCGGCTATTCGGAGGCTTACTTCGCGGAGTTCAAGTTCCAGGCC 843306
QY 616 ThrGluGluGluMetGluThrAla 623
DB 843305 TCGTCAGACACTTCGCAACAACGCTTATTCGAGGTGAGGTGATGCTCGCCGCGGACCG 843246
QY 624 624
DB 843245 GTGCGCGCGTTCGAAACCGGACGCTC 637
QY 637 Ile 637
DB 843191 GCGAACTCCCGCGGAAGGCGCTCGATCAACGCCCGACGATTCGACCGCTCGCG 843132
QY 638 AlaArgMetValGlySerLysThrValSerGlnCysLysAsnPheThrPheAsnTyrLys 657
DB 843131 GCCCGCGCTTGGCGGACGAGATTTTCA 843102
QY 658 LysArgGlnAsnLysAspGluIleLeu 669
```

```
DB 843101 -----TCGAGATTGACACCGCTGTGCGCATGATGACACCGGTAAACGATACACAGCAT 843048
QY 670 LysLeuLysMetGluLysGluArgAsnAlaArgLysLysLysAlaProAla 689
DB 843047 CATCTGCGGTC-GACCGCGCGGAAGGATGCTGTCAACGCGCGCGGCTCGACGAGCGCGC 842989
QY 690 AlaSerGluGluAlaAlaPheProValValGluAspGluGluMetGluAlaSerGly 709
DB 842988 CGCTACATCAAGCCGTTCCAGCTTCGACCGGACGACGACCGCGCGCGCGCGCGCGC 842929
QY 710 ValSerGlyAsnGluGluMetValGluGluAlaLysAlaLysSerGlyAsn 729
DB 842928 GCGCGTGGG----- 842920
QY 730 GluValProArgGlyGluCysSer-----GlyProAlaThrValAsnAsnSer 745
DB 842919 -----GCGCGGTGACACCGCGGATTCGCGGCTTACCGCTTACCGCGCT 842872
QY 746 SerAspThrGluSerIleProSerProHisThrGluAlaAla----- 759
DB 842871 GCGCGCGTTCGCGGCTTCGATGACAGACGCGCGCTGCGCGGACGCGCGCTTCGCGC 842812
QY 760 -----LysAspThrGlyGlnAsnGly-----ProLysProProAlaThrLysGlyAl 775
DB 842811 GATGTGAAAGCACCGGACAAACCGGTGCGCGCGCGCGCGCGCGCTTCGAGACC 842752
QY 775 AlaSerPro-----ProProGlyProProThrProPro-----ArgArgThrSerAr 791
DB 842751 GAACTCTCGGACTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 842692
QY 791 GAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProProProAl 811
DB 842691 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 842635
QY 811 aProProSerProSerAlaPro-----ProProValaProLysGluGluLys 827
DB 842634 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 842575
QY 827 aGluGluGluThrAlaAlaAlaProProValaGluGluGluGluGluLysProProAl 847
DB 842574 GCGGAAGACAGACCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 842539
QY 847 AlaGluGluLysAlaValaAlaPheThrGlyLysAlaGluGluProValaLysSerGluCysTh 867
DB 842538 GACCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 842506
QY 867 rGluGluAlaGluGluGluProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAl 887
DB 842505 -----GCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 842453
QY 887 aGluGluAlaLysLysAlaGluLysGluGluGluGluGluGluGluGluGluGluGluGlu 907
DB 842452 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 842393
QY 907 sSerSerGlyAlaProGln-----AspSerAspSer 917
DB 842392 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 842333
QY 917 rSerAlaThrCysSerAlaAspGluValaAspGluAlaGluGluGluGluGluGluGluGlu 937
DB 842332 CGTTCACACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 842305
QY 937 uLeuSerProArgProSerLeuLeuThrProThrGlyAsp-ProArgAlaAsnAlaSerP 957
DB 842304 -----AGCCCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 842270
QY 957 rGluLysPheProLysPheLysGluLysGluLysGluLysGluLysGluLysGluLysGluLys 977
DB 842269 CCGCTCCCGCGCGTGG-----GCGAACACCGCGCGCGTGA- 842238
QY 977 InValThrLysValHisGluProProArgGluAspAlaAlaLysProThrLysProAlaPro 997
```

Db 842237 -----CCGCGCGGCGCCGCGCATTCGACGCGCGCGCAC 842198
 Qy 997 roAlaProProProGlnAsnLeuGlnProGluSerAaAlaProGlnGlnProGlyS 1017
 Db 842197 CCGCGCGCGCGG-----GCCCGCGCGCGCGCCCAAGA 842168
 Qy 1017 eGSeProAArgGlyLysSerArgSerProAlaProAlaAaPlyGluAlaPheAla 1037
 Db 842167 GCGCACCG-----GCGGTGCGCGCGCGCGCATTCGACGCGCGCGCGCAC 842117
 Qy 1037 laGluAlaGlnLysLeuProGlyAaPProProCyStrPThrSerGlyLeuProPhePro 1057
 Db 842116 CACCGCTTCGCGCGCGCGCGCGCGAGCAAGCAAGCGCGCTCCGCGCGCGCGCAA 842057
 Qy 1057 aLProProAArgGluValIleLysAlaSerProHisAlaProAaPProSerAlaPheSer 1077
 Db 842056 TCCGCGCGAGGCGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 841997
 Qy 1077 yAlaProProGlyLysPProLeuProLeuGly-----LeuHisAaPThrAlaAaPPro 1095
 Db 841996 AGATCCCG 841937
 Qy 1095 aLeuProAaPProProThrIleSerAaPProProLeuLleSerSerAlaLysHisP 1115
 Db 841936 GCCCGCGGTTCGCGCGGTGCGCGCAACCGCGCGCGCGCGCGCGCGCGCGCGTCC 841877
 Qy 1115 roSerValLeuGlnAArgGlnIle----- 1122
 Db 841876 CCGGAGTCCGCTTGGCGCGCGCGCGCGCATCGACGCGCGCGGTGAGCGGTGAGAGGGT 841817
 Qy 1123 -----GlyAlaIleSerGlnIyM 1129
 Db 841816 CGTTACCATATTGATCATCGTCTCGTCAGAGGTGTCAGAGGCGAGGTCTGGCGGG- 841758
 Qy 1129 eSerValGlnLeuHisValProTySerGlnHisAlaLysAlaProValGlyProValT 1149
 Db 841757 --GGCGTAAAGCGGTCCAGCGCGCCCAACGAGCGCGCGGTACCGCGATCGCGCGCTTGC 841700
 Qy 1149 hMetGlyLeuProLeuProMetAaPProLysLysLysAlaProPheSerGlyValLysG 1169
 Db 841699 CAGGGTCCGCGCGGTCCCGCGCGCGCG-----CGGTGCGCGCGGTGCGCA 841652
 Qy 1169 InGlnLeuSerProAArgGlyGlnAlaGlyProProGluSerLeuGlyValProThz 1189
 Db 841651 TCAACAGCGCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 841604
 Qy 1189 la---GlnGluAlaSerValLeuAArgGlyThraLaleuGlySerValProGlyLysert 1208
 Db 841603 CCGGTCTCGCGCGCGGTGCGCGCGGTGCGCGCGGTGCGCGATCAACCGCGCGCGCGCG 841544
 Qy 1208 leThrLys-----GlyIleProSerThrArgValProSerAaPProAlaIleThrt 1225
 Db 841543 CCGGCGAG 841484
 Qy 1225 yArGlySerIleThrHisGlyThrProAlaAaPValLeuTyLysGlyThrIleThz 1245
 Db 841483 CCGCGCGAGCGCAAGAACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 841452
 Qy 1245 rgIleIleGlyLysAaPProSerProSerArgLeuAaPArgGlyAArgGluAaPProL 1265
 Db 841451 -----CGCGGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 841397
 Qy 1265 yGlyLysIleValIleTyArgGlnGlyLysLysLysLysLysLysLysLysLysLys 1285
 Db 841396 CCGGCG 841353
 Qy 1285 eValThrGlnCysSerLysGluAaPArgLysSerSerGlyProProHisGluThz 1305
 Db 841352 -----GCGTTG 841325
 Qy 1305 laAlaProLysAaPThrTyArgPheMetGlnGluArgValGlyAlaAlaIleSerSer 1325
 Db 841324 CCGGACCG-----GCGCGCTG 841310

Qy 1325 laSerIleGlnGlyLeuMetGlyArgAlaIleProProGluArgHisSerProHisL 1345
 Db 841309 CCGCGTTCCTCCCGCG--CGCGAAACAGCGCGCGCGCGCGCGCG-----CGCGCGCTGC 841257
 Qy 1345 euLysGlnGlnHisValIleArgGlySerIleThrGlnGlyIleProAArgSerTyValG 1365
 Db 841256 CCGCGCGGTGCGCGCG-----CGTTCCAGAAATTC 841227
 Qy 1365 luAlaGlnGlnAaPThrLysAArgGluAlaLysLeuLysLysAArgGlnGlyThrPPro 1385
 Db 841226 GCGGAGCG 841168
 Qy 1385 ro---ProProProProSerArgAaPLeuThrGlnAlaTyLysThrGlnAlaLys 1402
 Db 841167 CCG 841108
 Qy 1403 -----GlyProLeuLysLeuAaPProAlaHisGlnGlyLeuValAlaThrv 1418
 Db 841107 ACAGCG 841070
 Qy 1418 aLysGlnAlaGlyAArgSerIleHisGlnLleProAArgGlnGlnLysAaPProG 1438
 Db 841069 -----CGCGGCG 841018
 Qy 1438 luLeuProLeuAlaProAArgProLeuLysGlnGlySerIleThrGlnGlyThrProLeu 1457
 Db 841017 GCGCGCGGTGGCTCGCGGTCCG-----GCGCGCGCGGTGG 840982
 Qy 1457 ----- 1457
 Db 840981 CCGCGTTCGCGATCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 840922
 Qy 1458 --LysTyAaPThrArgGlyAlaSerThrThrGlySer--LysLysHisAaPValAaPProL 1476
 Db 840921 GCG 840862
 Qy 1477 IleGlySerProGlyArgThrPheProProValHisProLeuAaPValMetAlaAaPAla 1496
 Db 840861 CCGGACGAACTGGCGGTAAACCGCGCG-----CGTCTG----- 840826
 Qy 1497 ArgAlaLeuGlnAArgLacyTyArgGlnGlnSerLeuLysSerAaPProGly 1513
 Db 840825 -----CGTCTGCGCGCGCGGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 840793
 Qy 1514 ---ThrAlaSerSerGlySerIleAlaArgGlyAlaProValIleValProGlu 1532
 Db 840792 CGAACAGCG 840733
 Qy 1533 LeuGlyLysProAArgGlnSerProLeuThrTyGluAaPHisGlyAlaProPheAlaGly 1552
 Db 840732 TGTTGACGCGCG-----CGCGCG 840715
 Qy 1553 HisLeuPro-----ArgGlySerProValThrMetAArgGlnProThrvPro 1567
 Db 840714 CATTAACCGAGGTAATCG 840655
 Qy 1568 ArgLeuGlnGlnGlySerLeuSerSerSerLysAlaSerGlnAaPArgLysLysLys 1587
 Db 840654 GCTCCGCGAAGCATCAACCATGACATGACATGACATGACATGACATGACATGACATG 840601
 Qy 1588 -----ThrProAArgGlnAla 1593
 Db 840600 TGGTGCG 840541
 Qy 1594 -----LysAaPProHisSerThrValProGlnHisAaPProHisAaPProL 1610
 Db 840540 TATCGGATCCG 840490
 Qy 1611 TyArgLysLysLeuAaPArgGlyValAaPArgGlyValAaPArgGlyValAaPArgGly 1629
 Db 840489 GCTGCTGACCG 840430

QY 1630 -----LeuAlaPheAspProThrSerIleProArgIleProLeuAspAlaAla 1647
 DB 840429 ACAGGCGCGCGCGCAACCCGCA-----CCGCCCGATAGCCGTCGACAGACGCC 840376
 QY 1648 AlaTYrTYrLeuProArgHisLeuAlaProAsnProThrTYrProHisLeuTYr- ProPr 1667
 DB 840375 GCG-----ACGGGCGCTTGACAAAGCGCTCGAGCGCCGGGAGTCTTCGCGCG 840328
 QY 1667 oTYrLeuIleArgIleTYrProAspThrAlaAlaLeuGluAsnArgIleThrIleAs 1687
 DB 840327 ATACTGGGCGCGCG- GCCCCCATTTGGCGCGCC----- 840292
 QY 1687 nAsPTYrIleThrSerGlnIleMetHisIleAsnThrAlaThrAlaMetAlaGlnArgAl 1707
 DB 840291 -----TCCGAGCCGATCTGCATCCAGCGCGAGTGTAGCGCCGCAAGCCAG 840242
 QY 1707 aAspMetLeuAlaArgIleuSerProArgIleuSerIleuAlaLeuAsnTYrAlaAla 1727
 DB 840241 CCGTGCCTCCGCGCTACGCGAACCCCGTTGAGCCGTTTGGCCGACAGCAACCCGCGG 840182
 QY 1727 Y-Pro-----ArgGlyIleIleAspLeuSerGlnAlaProHisLeuProValLeuValP 1745
 DB 840181 CCGCGCGCGGTTCGGGGGCTGCGCGGCTTGCGCGCGGTGCGCGCGGTGCGCGGTTG---C 840125
 QY 1745 rProThrProGlyThrProAlaThrAlaMetAlaAspArgLeuAlaTYrLeuProThrAlaP 1765
 DB 840124 CCGCGGTGCGATCGACCGCGCGCTGCG-----CCGACGCCAC 840086
 QY 1765 rGlnIleProPheSerSerAlaHisSerSerSerProLeuSerProGlyGlyProThr-His 1784
 DB 840085 CCGCGCG-----CCGAAGCCAGCCAGCCGCGCATTCGCGCGGCCACGACCCGAC 840032
 QY 1785 -----LeuThrIleuProThr-----ThrHisSerSerSerGlu 1795
 DB 840031 CCGCGGTGCGCAACCCGCGCAACCCGCGCGCGCGCGCGCTGAGCAACAGACGCCGCGCT 839972
 QY 1796 ArgGluArgAspArgAspArgGluArgAsp-----ArgAspArgGluArgGluys 1812
 DB 839971 TGCACCGCGCGCGCGCTGCGCGCGCGCAAGAAACCAATCCGCGCGCGCGCGCGCGC 839912
 QY 1813 SerIleLeuThrSerThrThrThrValGlu-----HisAlaProIleThrPargProGly 1830
 DB 839911 CCGTGCAGAACAGACCCCTGCGCTACCGCGCGCGCGCGCGCTGCCACCGGTGTCAGGT 839852
 QY 1831 ThrGlnGlu-SerSerGlySerSerGlySerSer-----GlyGlyGlyGlyGlySerse 1848
 DB 839851 TTTCGCGGTACCGCGCGGCTCGCGCGCGCGCGCGCGCGCGGAGCATTCGCGCATTC 839792
 QY 1848 rSerArgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrG 1868
 DB 839791 CCGCGCGCGCGCAACCCGCGCTGCGCATTTAGTGTGAGCGCGCGGTCCGCGGTGAC 839732
 QY 1868 nAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetGlyIleIleTh 1888
 DB 839731 CTCCGCGCGACCGATCAGCGCGC---CAGCATGCTCGCGGTCCGCGCGC----- 839683
 QY 1888 rAlaValGluProSerIleuProThrValLeuArgSerThrSerProValSerSer- ProVal 1908
 DB 839682 -----CCACCGCGCGCGCGCTC-----GCCACCAACAGCGCGGTTC 839645
 QY 1908 rGPProAlaAlaThrPheProProAlaThrHisCysAspProLeuGlyIleThrLeuAspGlyV 1928
 DB 839644 CACCG 839607
 QY 1928 aTYrProThrIleuMetGluProValLeuLeuProGlyGluAlaProGValAlaArgP 1948
 DB 839606 -----CCGACCCCGCTCCCGCGCGGTGCGCGCGGT-----ACGCTCTGCTC 839561
 QY 1948 rGluIleArgProArgAlaAspThrGlyHisAlaPheLeuAlaLeuProProAlaArgSer 1968
 DB 839560 CCGCGATCCCG 839507
 QY 1968 TyLeuGluProAlaSerSerProSerIleuGlySerGluProArgProLeuValProPro 1988

DB 839506 CC---CAACCG 839450
 QY 1988 aSerGlyHisAlaThrIleAlaArgThrProAlaLeuAsnLeuAlaProHisAlaAs 2008
 DB 839449 TACCC---CAACGCGCG---GACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 839396
 QY 2008 ePProAspProProAlaProProAlaSerAlaSerAspProHisValGlyIleThrGln 2028
 DB 839395 CCGCGGACCG 839355
 QY 2028 eTyPProPheSerIleGlnGluLeuGluLeuArgSerIleuGlyTYrHisGlySerSer 2048
 DB 839355 ----- 839355
 QY 2048 YrSerProGluGluValGluProValSerProValSerSerProSerIleuThrHisAsp 2068
 DB 839354 -----CCGGGGCGCGCGGTGCG-----CGCGTGGCGCGCGCGCG 839321
 QY 2068 ySGlyLeuProIleHisLeu-GluGluLeuAspIleSerHisLeuGluGlyIleuAsp 2087
 DB 839320 CCGGGGATGCG-----TTGCGGATCAGCGGTGCGCGCGGTCAGTCTTGCTGCGCGCT 839267
 QY 2088 rPolySGlnProGlyProValIleuGlyIleuGlyAlaAlaHisLeuProHisLeuArg 2107
 DB 839266 TGAATCGCGCGCGCGCGCGCGGT---GTTTCAAGGTGTGACGCGGAGAG-GCGCTGCGCGG 839211
 QY 2108 rProLeuProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValys 2127
 DB 839210 GCGTTGAACCGCTCG 839160
 QY 2128 GlnHisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTYr 2147
 DB 839159 -----GTATCACCGCGCGGTGCGCGCGGT----- 839136
 QY 2148 ThrArgHisIleProGlnGlnLeuSerAlaPProLeuProAlaProLeuTYrSerPhePro 2167
 DB 839135 -----CCGCGATGCGCGCGGTGCGCGCGGTGCGCGATGACAGACCCCGTGCGCGG 839085
 QY 2168 GlnAlaSerCyPProValLeuAspLeuArgArgProPro----- 2180
 DB 839084 GCCCGCGCGGTACCGCGGTACAGCGCGCTTCCACCGCGCGCGCGCGCGCGCGCGCTTA 839025
 QY 2181 -----SerAspLeuTYrIleuProProProAspHisGlyAlaProAlaArgIleSerPro 2198
 DB 839024 CCGATCAGCAACCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTTCGCG 838965
 QY 2199 HisSerGluGlyIleuAspArgSerProGluProAsnIleThrSerVal-----Leu 2215
 DB 838964 AACCG 838906
 QY 2216 GlyGlyGlyGluAspGlyIle---GluProValSerProProGluGlyMetThrGluPro 2234
 DB 838905 GCGCGCTCCCGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 838846
 QY 2235 GlnHisSerArgSerAlaValTYrProLeuLeuTYrArgAspGlyGluGlnThrGluPro 2254
 DB 838845 CAGCACACCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 838810
 QY 2255 Ser-----ArgMetGlySerIleYSerSerProGlyAsnThrSerGlnPro 2268
 DB 838809 ACCGTACACCTAGCGCGGTCCCGCGCGGTATCCACCGGTGCGCGCGCGCGCGCGCGCG 838750
 QY 2269 rProAlaPheSerIleuThrGluSerAsnSerAlaMetValYSerIleYsGln 2288
 DB 838749 CCGCGGTGCG 838717
 QY 2289 GlnIleAsnIleYLeuAsnThrHisAsnArgAsnGluProGlyTYrAsnIleSerGln 2308
 DB 838716 ---GTCGTTGAGAAACCGATCCCGCGGTCCCGCGCGCGCGCGCGCGCGCGCGCGATAG 838660
 QY 2309 ProGly-----ThrGluIlePheAsnMetProAlaIleThrGlyIleu--- 2324


```

Dh 838659 CCGCGGACGAGCCACCGAGCTCCGCGAGTCCCGGACCCGACCCGAGTACGGCGCGAG 8386600
Qy 2325 -----MetThrTyArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGly 2341
Dh 838599 CCGCGCGGCGCGCGCGCTGCCCGCGGACACCAACGACCGCGCGCGCCACCAAC----- 8385464
Qy 2342 LeuGluAlaIleGlnArgLysAlaLeuMetGlyLysTyThrAspGlnThrGluGluSerPro 2361
Dh 838546 ----- 8385464
Qy 2362 ProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuPro-----Ala 2379
Dh 838545 CCGCGCGGCGCCACCGCGCGCGCGCGCGCGGAGGCCCGCGGACCCCGCGGACCCGCGGCTGCC 8384864
Qy 2380 AlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeu----- 2394
Dh 838485 CAACAACGAGCCACCTGACCCACCGGCAACCCCGCGCGCGCGCATCACTACCTGGAACAT 8384264
Qy 2395 -----ThSerProGlyGly-----GlyGlyLysAlaLysValSer 2406
Dh 838425 CCGCGCGGCGCCACCAAGCCCGCGCGGTGCCGAACAACCGCGCGCGCGCGCGCGCGCGCC 8383664
Qy 2407 -GlyArgProSerSerArgLysAlaLysSerProAlaProGlyLysAlaSerGlyAspAr 2426
Dh 838365 GGGCAGGCGCCGAGCCACGCGCGCGCGGACCCCGCGCGCCACCTTACGAGC---AACCA 8383094
Qy 2426 GProProSerValSerSerValHisSerGluGlyAspCyAsnArgArgThrProLeuTh 2446
Dh 838308 CCGCGCGCGCGCACCATTTGGCCCCCGTGC CGGG-----GCCCGTTGGC 8382644
Qy 2446 rAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPhe---ProTyrAs 2465
Dh 838263 GCCGTGGCGATCAACGCGCGCGCGCGGTGAGCGGACACATTGGCATTGACGCGCGCCAG 8382044
Qy 2465 nProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProGlyLeuP 2485
Dh 838203 CAGGTCTCTGCTGACGCTCTGCGCGGGATTGGCGAGCACCGCGCGCGCGC -GGCTTCG 8381454
Qy 2485 alaGlySerGlyPro 2490
Dh 838144 CGCTGGCGTAGGCC 838129

RESULT 10
US-08-804-227C-7/c
; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kuhstose, Stuart A.
; APPLICANT: Rotheck, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII (DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231

```

```

: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 317-276-2459
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4437 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
:
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 350..14002
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 14046..20036
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 20110..31284
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 31329..36071
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 36155..41830
:
: US-08-804-227C-7
:
:
: Alignment Scores:
: Pred. No.: 5,22e-21 Length: ** 44377
: Score: 670.00 Matches: 665
: Percent Similarity: 31.22% Conservative: 282
: Best Local Similarity: 21.93% Mismatches: 1121
: Query Match: 5.07% Indels: 980
: DB: 2 Gaps: 149
:
: US-09-522-753-5 (1-2517) x US-08-804-227C-7 (1-44377)
:
: QY 9 A|a|c|n|t|t|p|a|g|a|t|t|n|g|l|u|p|r|o|f|h|i|s|e|r|l|e|u|s|e|r|t|y|r|p|o|v|a|l 28
: Db 9254 G|C|C|C|G|T|A|G|T|G|C|C|G|A|G|C|C|G|C|G|T|G|A|T|C|C|C|A|G|C|C|G|C|C|G|T|A|C 9195
: QY 29 G|n|l|e|a|a|g|t|t|n|t|t|h|a|p|v|a|g|l|y|e|u|e|u|g|t|y|r|g|l|n|t|t|h|i|s|e|r|a|a|p 48
: Db 9194 G|n|t|c|c|g|c|t|c|a|c|a|c|a|t|----- 9177
: QY 49 T|y|r|a|s|e|r|h|i|s|e|r|p|r|o|g|i|l|e|g|i|n|p|r|o|g|i|n|a|r|g|a|r|g|p|r|o|s|e|r|u 68
: Db 9176 -----CC|A|G|T|G|G|G|G|G|A|C|C|A|G|C|C|G 9153
: QY 69 L|e|u|s|e|r|l|u|p|h|e|g|i|n|p|r|o|g|i|l|y|a|e|n|g|l|a|r|g|s|e|r|c|i|n|l|u|e|h|i|s|e|u|a|r|p|r|o|g|i|l|u|s|e|r 88
: Db 9152 C|A|T|C|A|G|C|C|A|G|C|C|T|C|A|G|G|A|T|C|A|C|C|G|C|G|T|G|C|G|G|A|C|C|A|C|T|G|C|G|C|C|T|C|A|G|C 9093
: QY 89 H|i|s|e|r|t|y|r|l|e|u|p|r|o|g|i|l|e|u|g|l|y|s|e|r|c|i|n|u|e|h|i|e|g|i|l|u|s|e|r|t|y|r|a|r|g|p|r|o 108
: Db 9092 C|A|T|T|G|C|-----T|G|C|A|C|G|T|C|T 9075
: QY 109 A|r|g|h|e|u|l|e|u|e|u|p|r|o|a|p|-----P|r|o|e|u|l|e|u|a|r|g|p|r|o|s|e|r 121
: Db 9074 G|A|-T|T|A|C|C|G|C|A|C|T|G|C|C|C|G|C|A|C|C|G|C|A|C|C|C|G|A|G|A|T|G|A|C|C|T|T|G|C|G|C|G|C|G|T|G 9016
: QY 122 P|r|o|e|u|e|u|a|l|a|r|h|g|i|n|-----P|r|o|a|l|g|l|y|s|e|r|g|i|u|a|p|l|e|u|h|t|l|y|r|a|a|r|g|s|e|r 140
: Db 9015 G|A|A|G|C|C|G|T|C|A|C|A|G|C|A|C|A|C|C|A|C|A|C|C|T|C|C|C|A|C|G|C|G|T|C|C|G|T|C|-G|C|A|C 8957
: QY 141 L|e|u|h|t|t|g|l|y|s|e|u|g|i|u|p|r|o|v|a|l|s|e|r|p|r|o|p|r|e|r|p|r|o|p|r|o|----- 154
: Db 8956 C|T|C|G|G|A|A|G|C|C|T|T|G|A|C|G|C|C|G|T|C|G|C|C|G|C|C|A|G|C|C|C|C|G|T|G|C|G|C|G|A|A|T|T|C|A|C 8897
: QY 155 -----H|i|s|t|h|A|a|p|P|r|o|-----G|i|u|e|g|i|u|l|e|u|a|l 163
: Db 8896 G|A|A|C|A|C|C|G|T|G|C|G|T|G|C|C|A|T|C|A|C|G|T|C|A|C|C|C|A|C|G|C|C|A|A|G|C|C|A|G|A|T|C|A|C|A|C|T|A|C 8837
: QY 164 P|r|o|r|o|a|r|g|l|e|u|s|e|r|t|y|r|g|i|u|g|i|u|l|e|g|i|n|a|n|t|a|a|r|g|v|a|l|a|a|r|g|v|u|l|e 183

```


QY 763 GtYglnAbnglyProlys-----ProProAlaThrleuGlnAlaAspGly 777
|||:::|||||
Db 6868 GGGGAGACA-GCGCCCGAGCAAGTACGAGACACCGGGGCGAGCATCTCCGGGAGAGT 6810
QY 778 ProProProGlyProProThrProPro-----ArgArgThrSerArgAlaProIle 794
|||||:::|||||
Db 6809 CCGCCCAAGCGCGCCGAGCGCGCGCATTCGAGCGCGGAGCATCTCGGATCCGACCA 6750
QY 795 GluProThrProAlaSerGlnAlaThr---GlyAlaProThrProProPro----- 810
|||:::|||||
Db 6749 GCTGAGCGCCCGAGCAAGAGGGGTCTCGTCGCGACGACGCGCGCTCAGCGGTGG 6690
QY 811 -----AlaProProSerProSer----- 816
|||:::|||||
Db 6689 TGATCCAGAGCGGGGCTTGGGGTACCGCTCGCCGACGCGCTGAGACGAGGCGACGCGCT 6630
QY 817 AlaProProProValProlysglnGlnLysGlnGlnGlnAlaAlaAlaProPro 836
|||:::|||||
Db 6629 CCGCCCGCTCCGGCTCGCGCGGGGCGAGCCACAGAACT--CGGGTGTGCGGGGCGGACCG 6571
QY 837 ValGlnGlnGlnGlnGlnGlnLysProProAlaAlaGlnGlnLysAlaAlaAspThrGly 856
|||:::|||||
Db 6570 TCACCGCGCGCGCGCGAGAGTCCCGGCTGCTCCCG-----CCTGTACCGACCGCGC 6517
QY 857 LysAlaGlnGlnProValLysSerGlnGlyThrGlnGlnGlnGlnGlnGlnGlnGln 876
|||:::|||||
Db 6516 ACGGCG-----ACGGTGTGAGCTCGCGCGCGCGCGCGCGAGTTCAGCTGGAGTCCG 6463
QY 877 GtYlYsAbpAlaGlnAlaAlaGlnAlaThrAlaGlnGlnAlaLeuLysAlaGlnLys 896
|||:::|||||
Db 6462 GGGGCGAGTCCGGAGTCCGGC---GTGTCTCTCGTCGGGGCGAGAGACACATGACCG 6406
QY 897 GlnGly---GtYserGlnArgAlaThrThrAla----- 906
|||:::|||||
Db 6405 GAGAGCGCGGGTTCGGCGGTGACGACGCGCGGTGCCAGCGAGCGGTAGCGGATC 6346
QY 907 -----LysSerSerGlnAlaProGlnAsp 914
|||:::|||||
Db 6345 CCGGCGGTGTGACCGCTCCCGCTCAGTGGCCCGAGCGTCCGCGGAGCGCGGGGCG 6286
QY 915 Ser----- 915
|||:::|||||
Db 6285 TCGAGCGCGGTGACGCGGTGCGAGCGGGGTCCAGCGAGTACGGGTGCGGTGAAGGG 6226
QY 916 -----AspSerSerAlaThrCys 921
|||:::|||||
Db 6225 TAGGTGGGAGGTGATGCGCGCGCGGTGCTCCGGTGTGAGTCCGCGGTGCGTCCAG 6166
QY 922 SerAlaAspGlnValAlaAspGlnAlaGlnGlnGlnGlnGlnGlnGlnGlnGln 941
|||:::|||||
Db 6165 CCGGCGTCCCGGGTGAACACCGTCCGAGCGCGCGCTCGAACCTTCGCGCTCCGACGG 6106
QY 942 ProSerLeuLeu-----ThrProThrGlnAspProAlaGlnAlaAspProGln 958
|||:::|||||
Db 6105 CCGCGCGGATGCGCGGACACACCATATCATCCGCGCGCGAGGACCTACCGCG- 6047
QY 959 LysPro-----LeuAspLeuLysGlnLysLysGlnLysGlnAlaAla 972
|||:::|||||
Db 6046 CATCCCGGTGAGACCCCATGCGGACCCACCTTCACCAACGTCCTCCGACCCGAGACCG 5987
QY 973 IleProProIleGlnValThrLysValHisGlnProProAlaGlnAlaAla-----Ala 990
|||:::|||||
Db 5986 CACCAACCCCAACCCCTCCGCGAAGACGACCG--CTCTCGCGCATGACGACCAATACC 5928
QY 991 ProThrLysProAlaProProAlaPro-----ProProPro 1002
|||:::|||||
Db 5927 CCGGATCCCGCAACTCCCGAAGACACGACCCGACCGACACCGGACCGACCGACCA 5868
QY 1003 GlnAsnLeuGlnProLys-----SerAlaAlaProGlnGlnProGlnLysSerProArg 1020
|||:::|||||
Db 5867 CACCGCGCGGACCGGAGAACTCCACGAGTCCACCAACCCCGGAGACTCCCA--- 5811

QY 1021 GtYlYsSerArgSerProAlaProProAlaAspLysGlnAlaAlaGlnAlaGln 1040
|||:::|||||
Db 5810 -----ACACCGGCTCATGACACCGAATGAACCCATGG-----ACACAT 5769
QY 1041 LysLeuProGlnLysProProCysThrThrSerGlnLysLeuProProValProProArg 1060
|||:::|||||
Db 5768 CCAACCGCGCGGCTCATCCCATCCCATCCATGAGGAGGAGGACACCGATCTCA----- 5715
QY 1061 GlnValIleLysAlaSerProHisAlaPro---AspProSerAlaPheSerTYAlaPro 1079
|||:::|||||
Db 5714 -----GCAACCCACATCACCGGACACACACCGACCGGCGGCTTCACCG 5667
QY 1080 Pro-Gly---HisProLeuProLeuGlnLys-----HisAspThrAlaAsp 1093
|||:::|||||
Db 5666 CCGCAACCGGACACCTCCCGACCCCTCAACACCCCGGACACCGATCTCGAGC 5607
QY 1093 gProValLeuProArgProProThrIleSerAspPro----- 1105
|||:::|||||
Db 5606 CCCCACCGACCATCCCGCCACCGCGCAACCCCATCAACCCGCGCGCGCA 5547
QY 1106 -----ProProLeuIleSerSerAlaLysHisProSerValLeuGlnArgGlnIleG 1123
|||:::|||||
Db 5546 CCAACCAACCGGACCGGATGACACGACGACACCCCGGACATGACGACGAC----- 5492
QY 1123 yAlaIleSerGlnGlnMetSerValGlnLeuHisValProTYSerGlnHisAlaLysAl 1143
|||:::|||||
Db 5491 -----CACCTCCCGCACCG--ANTGACCAACACCG 5464
QY 1143 aProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaPr 1163
|||:::|||||
Db 5463 ACCGACACCTCCA-----CACCCGAGCTTCGCAACCCCGGAGCAACGCGCACCC 5416
QY 1163 oPheSerGlnValLysGlnGlnGlnLysSerProArgGlnAlaGlyPro----- 1180
|||:::|||||
Db 5415 TCCAGCGCAACAGCGCGAGCTGAGCAACTCGTTCGACCCGAC--AACCCACCCACAGA 5357
QY 1181 -----ProGlnSerLeuGlnValProThrAlaGlnAlaSerValLeuArgGlnTh 1198
|||:::|||||
Db 5356 ACCGAAACGAAACGACCCGACGACCCGACCGGACCGGATTCAGCCCGGACACCGCG 5297
QY 1198 rAlaLeuGlnSerValProGlnGlySerIleThrLysGlnIleProSerThrArgValPr 1218
|||:::|||||
Db 5296 GTCCACGTGAGGTCCGGAACATCATCCGCGCAACGACCGACCTCCCGCCAC 5237
QY 1218 oSerAspSer-----AlaIleThrTYAr 1226
|||:::|||||
Db 5236 CATCGAACACCTCATTCGACGACCTCGCAACACCCGCGCGGACCAACCCAGC 5177
QY 1226 gGtYserIleThrHisGlyThrProAlaAspValLeuTYLysGlyThrIleThrArgI 1246
|||:::|||||
Db 5176 CCCCATACCAACCCGATGCGGACCTGACCGGTG-----MACAAACCCCGAC- 5128
QY 1246 eIleGlnLysAspSerProSerArgLeuAspArgGlnArgGlnLysAspSerLeuProLysG 1266
|||:::|||||
Db 5127 -----CACCGCGGACGCGCGCGGACGCGACACCCCGAGACACCTCCCGGACG 5075
QY 1266 yHisValIleTYrGlnGlnLysLysGlnHisValLeuSerTYrGlnGlnGlnMetSerVal 1286
|||:::|||||
Db 5074 AAC----- 5071
QY 1286 lThrGlnCysSerLysGlnAspGlnYArgSerSerSerGlyProProHisGlnThrAlaAl 1306
|||:::|||||
Db 5070 -----CACCAAGACGACGACGAGCTTCGACGCGGACCAACGACGACGACCGC 5021
QY 1306 aProLysArgThrTYrAspMetMetGlnGlnArgValGlnArgAlaIleSerSerAlaSe 1326
|||:::|||||
Db 5020 A-----CGTGTCTCGAAAGAGTCCGCGTCCGAGC 4991
QY 1326 rIleGlnGlnLysMetGlnArgAlaIleProPro----- 1337
|||:::|||||
Db 4990 CAGCGAATAGCACAATCCCGGAGATCCGACCCGACCCGCTCCGATCTACACCAACCG 4931
QY 1338 -----GlnArgHisSerProHisHisLeuLysGlnGlnHisHisIleArgGlySe 1354

```

Db      4930 CACCGCTGCGCGCCCAAGCCCTCCGACCGCGCGCAACAACAGCGCGAGGAGC 4871
Qy      1354 rIlethrInglInyIleProArGserTyraGIuAlaInglInuArTyLeuArGI 1374
Db      4870 AGCGCGCTCCGGGTACCGGGTCAAGAGCAGCAGCCGTC- GATGCTTGAACCACT 4812
Qy      1374 uAlaIysleuLeu-----LysArGIuGI 1382
Db      4811 CCGGCGCTCTCGATGATGATGATGGCGGTGTCCCACTGATCCGGAACAGACAGCAG 4752
Qy      1382 yThrPro-----ProProProProProSerArg-----As 1392
Db      4751 CCGCGCGCGCGCGCGGTCCGACCGCGCGGACACCGCGCGCTCTGCGACAGGCGCA 4692
Qy      1392 pLeuThrGIuAlaTyTyLeThrGIuAlaLeuGIyProleuYsleuYsProAlaIngl 1412
Db      4691 CGGTGCGCGACGACGATGAGCTGCTGGCGGTGGCGGTCCGAGTCCGCGGCA 4632
Qy      1412 uGIyLeuValAlaThrValIysGIuAlaGIyArGIserIleInglIleProArGIuGI 1432
Db      4631 GCGTCCCGGTGCTCCAGGCGAAG-----CAGCA 4605
Qy      1432 uLeuArGIserThrProGIuLeuProleuAlaProArGIProleuYsGIuGIserIleTh 1452
Db      4604 TCTTGACCAACAC-----CGCGCGCGCGC-----GGCGCGCTG-- 4571
Qy      1452 rGIInglYThrProleuYsTyraSPhrGIyAlaSerThrInglYserIleYsIleAs 1472
Db      4570 -----GDTGACCGACGCTTGACCTTGACGAGCCCA 4539
Qy      1472 pValArGIserIleGIySerProGIyArGIrThrPheProProAlaInProleuArYa 1492
Db      4538 GCCACAGGGGTC-----GCCACCGCGTC----- 4514
Qy      1492 lMeAlaArGIAlaArGIAlaLeuGIuArGIaCySTyGIuGIuSerIleSerArgPr 1512
Db      4513 ---ACGCCGTACGTGGCGACGAGCGCGCTCGCTCGATCGGATGCCAGCGCGCTCC 4458
Qy      1512 oGIyThrAla-----Se 1516
Db      4457 CGGTACCGGTCCCGCTCCACCGCGGTCCAGCTCACCGCGCGAGCGCGGTGCGCAGGG 4398
Qy      1516 rSerSerGIyGIySerIleAlaArGIyAla-----Pr 1527
Db      4397 CAGCTCGATGACGCCCTCTCTGCGCGACCGGTGGAGACCGGTACAGCGCTTGCGCAC 4338
Qy      1527 oValIleVal--ProGIuLeuGIyYsProArGIInSerProleuThrTyGIuArGIhIs 1546
Db      4337 CGTCTGCTGTGACCGGACCTGCCCGCACCGCCGACGACCC----- 4296
Qy      1547 GIyAlaProPheAlaGIyIleuProArGIySerProValThrMetArg----- 1563
Db      4295 GGTGTCCCGACGCGCGGCGT--CGAAGACCGCTCACTGCTCACTACGAGCGACACCT 4239
Qy      1564 -----GIuProThrProArGIleuGIInglYserIleuSerSerSerIyAlaSerGIIn 1581
Db      4238 CGGACCGAGCCGTACCGTGGCATGGCGGAGAGACGATGCAAGCGCGCGGTGGGGGCA 4179
Qy      1582 AsPArGIySleuThrSerThrProArG-----GIuIleAlaYsSerProIleSerThr 1599
Db      4178 GCCCGCGCTGCGCGAGAACTCCACGAAATGCGCGGCGCCGACATCACCGTGGCGCGC 4119
Qy      1600 ValProGIuInIleArGIhIsProIleSerPro----- 1610
Db      4118 CCGCCAGAGCC-AGCCCGAC-----TCGCGCGCGCGACGCGCTGCAAGCGAGGTGC 4066
Qy      1611 -----TyrGIuInIleuLeuArGIyAlaSerGIyAlaSPleuTyraArg 1625
Db      4065 AGCGGCACAGGAGGAGGAGGACGCGGGTGTGACGCTGAGC----- 4024
Qy      1626 SerHIleProleuAlaPheArGIProThrSerIleProArG-----GIyIleProleu 1643

```

```

Db      4023 -----GCCGTCCGCG-----GTCCGAGGCTGTAGCCATCCGCGC 3985
Qy      1644 AsPAlaAlaAlaTyTyLeuProArGIhIsleuAlaProAsnProThrTyraProhIs 1663
Db      3984 GAGGCGAGCTTCGCG-----GTGGTCCGGGTGAGGAGTACCCCTGC 3943
Qy      1664 LeuTyPro-----ProTyLeuIleArGIyTyPro-- 1674
Db      3942 ---TATCCGTCCGCTCCCTGTGAAGCTCGGCGCTGACTCGGTGCGGTGCGCGCAG 3886
Qy      1675 AsPThrAlaAlaLeuGIuAsnArGIInThrIleIleAsnArTyIleThrSerGIIn 1694
Db      3885 AACACTCCGGCGTCCCGCGCGCGAGTGACTCGGGAGACATCGCGCTCTGAGGGCC 3826
Qy      1695 MetHisIleAsnThrAlaThr-----AlaMetAlaInArGIAla 1707
Db      3825 TCCACAGACGCTTCAGACGAGAAAGCGCTGCGGCTCATATGCGCTGCGCGCGC 3766
Qy      1708 AsPMetLeuArGIyLeuSerProArGIuSerSerIleuAlaLeuArGIAlaAlaGIy 1727
Db      3765 GATATC-----CCGAAGAACTCC-----GCGTCGAACCTCGCGCGC-- 3730
Qy      1728 ProArGIyIleIleAsPleuSerGIInAlaProhIsleuProValIleuValProProThr 1747
Db      3729 -----GAGTCAAGAAACCGCGCTCCGACATAGCTTGTCCGGCGCTC 3685
Qy      1748 ProGIy-----ThrProAlaThrAlaMetAsPArGIleuAla-----Tyr 1760
Db      3684 CCGGCGTGGGGTGTGACAGCCCGTTCGAGGTCCAGCGCGGTGCGGTGGAGAGCAG 3625
Qy      1761 LeuProThrAlaProGIuProPheSerSerArGIhIsSerSerProleuSerProGIy 1780
Db      3624 ATGGCGTGGCGCGCGACTCCAGCAGCTCCAGAAAGCGCGCGAGTCAAGCGCGCG 3565
Qy      1781 GIyProThrInIleuThrIleYsProThrInThrIleSerSerGIuArGIuArGIaPArG 1800
Db      3564 GGGAGCCCGGACCGACA---CCACAGATGCGAGTCCGCTCGCTCCGCGCTCCG 3508
Qy      1801 AsPArGIuArGIaAsPArGIaArGIuArGIuYsSerIleLeuThrSerThrThr 1820
Db      3507 CCGCGGGCGTCCGACACCCCTACCGGGCGACAGAGACATCGGGCCACCGGACAC 3448
Qy      1821 ValGIuInIleAlaPro-----IleTrArGIProGIyThrGIuInSerSerGIy 1836
Db      3447 GAGTCTTCGAGCCCGCGCGCTTCACGATGTCTCGGACGAGCAGCGAGGTGGGC 3388
Qy      1837 SerSerGIySerSerGIyGIyGIyGIyGIySerSerSerArgProAlaSerIleSerIle 1856
Db      3387 AGGTGTGAGACGAGGTGGCGGCGAGGGGAGCCCGGAGCGCTTCACAGCTGCAC 3328
Qy      1857 -AlaInglInIleSerProIleSerPro-----ArgThrGI 1868
Db      3327 AGCTGACCGCGGTACCG--AGTCAATGCCCTGGAGGTGAAGTGCCTGCGAGCGCG 3271
Qy      1868 nArGIAlaLeuGIInIleArGIProSerValIleuInIleAsnThrGIyMetYsGIyIleIleTh 1888
Db      3270 ACCTGGCGGAGTCTGCTG----- 3252
Qy      1888 rAlaValAlaGIuProSerIyProThrValIleuArGIser-----ThrSe 1902
Db      3251 -----GGCCAGAGCGCGCGGTGCTCTGCGGACCGAGCGAGCGCATCTCGGG 3199
Qy      1902 rThrSerSerProValArGIProAlaAlaThrPheProProAlaThrInIleCyProleuGI 1922
Db      3198 GACATCGCTCCGGTGAAGCGGAGT-----CGGCTGTGCGTGTGTCCTGTGG 3148
Qy      1922 yGIyThrIleuAsPArGIyAlaTyProThrIleuMetGIuProValIleuLeuProIyGIuAl 1942
Db      3147 GAGTTC-----CTTCGAGTCTC 3130
Qy      1942 aProArGIValAlaArGIProGIuArGIProArGIAlaSPThrGIyInIleAlaPheLeuAlaY 1962
Db      3129 TCCGAGTCCGGGAGACGCTGTGCGCCACG-----CATGGGTGGCGGTCA 3082

```


QY 393 -----AlaValIlePro-----PrometleuTyraAspAl 402
 Db 8108 GGTACGCAAGTTCAGAAACGCGTAGGAGAGTCGAACCCGCGGACCG-----GA 8061
 QY 402 aAaPGLInglInaTgIlelybPheIleAmeAaMgIleuMeAlaAspPro-----Me 420
 Db 8060 AGGCAAGTCCGGGTGACCGCGCTCGGC-----CGAGGCGTGGCGAGATCCCGCGCGCT 8004
 QY 420 tLveValTyrlYbAaPbAaPglInValMetAaMMeTTrpSerGluInglInuylbGluThrPh 440
 Db 8003 CCGAGCGCACAG-----CGTAGCAGAGACGACGCGCGT 7968
 QY 440 eAaPGLInuyls----- 443
 Db 7967 CCGCGGAGAGGCGCGACGAGACCGCGCGCTCGCCCGCTCGTCCGCTCC 7908
 QY 444 -----PheMetGlnIbProLybAaPbGlyLeuIleAaSerPheLeuGlu- 459
 Db 7907 CGCCCGGTGCGCGCTCGCGCGCGCTCGGTGAGAGCGGACGCTCGCGCG 7848
 QY 460 -----ArglybThr-ValAaGluCybValLeuTyrlTyrlLeuThrlybL 475
 Db 7847 CAGGCGCGGACGCGCGCGAGCGCGCGAG----- 7816
 QY 475 ysaAaMgIuAaenTyrlYbSerleuValArgArGserTyrlArgArGlylybSerGlnG 495
 Db 7815 --ACGCGGTGCAAGAGGACCGAGTACGCGCGGTGACCGAAGCGCGACGAGGTCTCC 7758
 QY 495 lngInglInglInglInglInglInglIn-----GlnGlnGlnG 509
 Db 7757 AGTTCACGTGCGCGACGACGAGGCTCTCGCGCGCGCGCTCGCTCCGCGCAGAG 7698
 QY 509 lngInPromeProArGserSerGlnGluInuylbAaPglInuylbGluylbGlu 529
 Db 7697 TCCCGCGCGCGCGCACCGCTGTCGAGAGCGTACGCGCGCGCTCGGCGTCAAGGCGG 7638
 QY 529 laGluylbGluInglInuyls-----ProGlu----- 537
 Db 7637 TGAACCGCGTCCCGTTCGAAGACTTCGCGCGCTTCGAGAGGCAATTCGCGACCGGCGC 7578
 QY 538 -----ValGluAaMgIuAaPblybGluAaPblybGluylbThrAaPbT 553
 Db 7577 ACGGCGCGCGCGAGCGATCGAGCGCGGCGCTCCGCGCGCGCGCTCGCGCGCG 7518
 QY 553 hrSerGly-----GluA 557
 Db 7517 CGTCGAGCGCGCGCTTGCGCGCGCGTGAACGCGCTCGCTCGTCCGCGACGCGCGA 7458
 QY 557 sPaAaPbGluylbGluAlaValAaAaSerlybGlyArGlybThrAaAaSerGlnGlyA 577
 Db 7457 CGATGAGAGAGAGAGACGACGAGCGCTCGAGCGGAAGGTCCGCGTCAAGCTCTCCAGGT 7398
 QY 577 rglrGlybGlyArGlyleThrArGserMeAlaAaMgIuAlaAaSerGluAlaIleT 597
 Db 7397 GACGCG-----GCGCGCGTGA 7383
 QY 597 hrProGlnInSerAlaGluLeuAlaSerMetGluLeuAaMgIuSerSerArGTrpThrG 617
 Db 7382 CTTGTTCGCGCAGTACGCGCTCCACCGCGTGGAGGTGAGAGTCCAGC-----ACAC 7329
 QY 617 lngInGluMetGluThrAlaLyblybGlyLeuLeuGluInuylbGlyArGTrpSerAlaI 637
 Db 7328 CGTCGTGAGACACACCTCGCGCGCTGACGTACGCGGACGAGGCGCGCTCTCGCGGAGGC 7269
 QY 637 leaAaPbMetValGlySer-----LybThrValSerGlnCybLybAaPbPheTyrlP 654
 Db 7268 CCGCGAGAGAGCGCGCGCGAGTCCCGGTGCGGAGTCCGAGAGCTCGACGCGC-CACT 7210
 QY 654 heAaenTyrlYblybArGglInAaMgIuAaPglInuylleuGlnGln----- 668
 Db 7209 TCGGAACGAGCGCGCTGACTCTCTGACGAGTTCGCGCACTCGCGGCGCGCGCGCG 7150

QY 669 -----HisLybLeuIleMetGluylbGluArgAaMAlaArg-ArglybLybLyb 685
 Db 7149 CGCGCGCGCGACAG-----GAGCAGACGCGCGCGCGCGCGCGCGCGCGCGCG 7099
 QY 686 AlaProAlaAlaAaSerGluGluAlaAlaPheProProValValGlu-----AspGlu 703
 Db 7098 GCGACCTGCGCA-----CCAGGCGACCGGTGCGCGCGGTGATCATGATACGTCCGAG 7045
 QY 704 GluMetGluAlaSerGlyValSerGlyAaMgIuInglInuylbGluValGluAla 723
 Db 7044 GACCACCGGTCTGTCGCGACAGACCGACGTCCAGGTGTCCTCGGTGCTCTCCCGACG 6995
 QY 724 leuHisAlaSerGlyAaMgIuValProArgGlyCybCyb-----SerGlyProAlaThrVal 742
 Db 6984 CGCGTGGCGGGCGCGGTGAGCGGTGACGCGAGCGGACCGGAGCGGAGCGCGCGCGG- 6926
 QY 743 AaMgIuSerSerAaPbThrGluSerIleProSerProHsThrGluAlaAlaLybAspThr 762
 Db 6925 AACCAACGCGCGCGCATGAGACGCG-----GAGAGCGCGTACCGCGCAACTGCTCTCCCG 6869
 QY 763 GlyGlnAaMgIuProlys-----ProbAlaThrLeuGlyAlaAspGly 777
 Db 6868 GCGCAGCA-GCGCGCGACGACGATACGAGACCGCGCGCGAGCACTCCGCGCGCAGGT 6810
 QY 778 ProProProGlyProProThrProPro-----ArgArGTrpSerArGAlaProIle 794
 Db 6809 CCGCGAAGCA 6750
 QY 795 GluProThrProAlaSerGluAlaThr--GlyAlaProThrProPro----- 810
 Db 6749 GCTGAGCGCGCGCGAGAGGGGTCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGGTG 6690
 QY 811 -----AlaProProSerProSer----- 816
 Db 6689 TGATTCAGAGCGGCGCTTCGAGTACGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCT 6630
 QY 817 AlaProProProValValProLybGluGluGluGluGluGluGluGluGluGluGluGlu 836
 Db 6629 CGCGCGCGGTCTCGCGTGGCGGGCGAGCGACAGAACT-CGGGTGCTCGCGCGCGCGCG 6571
 QY 837 ValGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 856
 Db 6570 TCACCGCGCGCGCGCGCAAGGTCTCCGCGCGCTCGCGCG-----CCTGTAACGACCGCG 6517
 QY 857 LybAlaGluGluProValLybSerGluCybThrGluGluGluGluGluGluGluGluGlu 876
 Db 6516 ACAGCG-----ACGTGTGACCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6463
 QY 877 GlyLybAaPbAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLybAlaGluLyb 896
 Db 6462 GGGCGAGTCCGAGTCCGCG-----GTGTCTCTCGCGCGCGCGCGCGCGCGCGCGCGCG 6406
 QY 897 GluGly--GlySerGlyArGAlaThrAla----- 906
 Db 6405 GCAAGCGCGCGGTGCGCGCGTGCAGACGCGCGGTGCCAAGCGAGCGCGGTAGCGGATC 6346
 QY 907 -----LybSerSerGlyAlaProGlnAsp 914
 Db 6345 CGGCGGTCTGACCGTCTCCGCTGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6286
 QY 915 Ser----- 915
 Db 6285 TCACCGCGGTACGCGCGGTGCGAAGGGGTCCAGCAGTACGCGTTCGTTGGAAGGG 6226
 QY 916 -----AspSerSerAlaThrCyb 921
 Db 6225 TAGTGAGAGGTGATGCGCGCGCGGTCTCCCGGTGAGATCGCGGTGCGGTCCAGG 6166
 QY 922 SerAlaAspGluValAspGluAlaGluGlyGlyAspLybAaMgIuLeuSerProArg 941
 Db 6165 CGCGGTCTCCGAGTACACGTGCGCGAGCGCGCGCTCGAACAACCTCCGCTCCGACGG 6106
 QY 942 ProSerLeuLeu-----ThrProThrGlyAspProArGAlaAaMgIaSerProGln 958

QY	1564	-----GIuPProThrProArgLeuGlnGlnuGlySerLeuSerSerSerLeuAlaSerGln	1561
Db	4238	CGAGCAGCAGCGCTGATCCGTGGCATTCGGCGAGAAAGGATCAGAGGCGCTCGGGGGCGGA	4179
QY	1582	AspArgIyIeLeuTherSerThProArg-----GluIleAlaYSerProHiserThr	1599
Db	4178	GCCGGCGCTGGCGGAGAACTCCAGAACATCGCGGGCCCGCAATCACCGTGGCGCGCG	4119
QY	1600	ValProGluIhVhIhProHisProIleSerPro-----	1610
Db	4118	CCGGCCAGAGCC-AGCCCGCAC-----TCGGCCCGGCGCAGCGCCTGCACGGCAGGTGC	4066
QY	1611	-----TyrgLhIhIeLeuLeuArgIyValIserGlyValAspLeuTyArg	1625
Db	4065	AGCCGCCAGGAGCAGAGAGACACCGCGGTGCAGCGGTAGC-----	4024
QY	1626	SerHisVilIeProLeuAlaPheAspProThrSerIleProArg-----GlyIleProLeu	1643
Db	4023	-----GCCGTCCGCG-----GTCGGAGGGGTAGAGCATCGCGCG	3985
QY	1644	AspAlaIalAlaAlaIaTyTYrTYrLeuProArgHisIleuAlaPheAsnProThrTYrProHis	1663
Db	3984	GAGGCCACGCTCGG-----GTGGTCCGGTGAAGCAGGTACCCCTCG	3943
QY	1664	LeuTYrPro-----ProTYrLeuIleArgIyTYrPro---	1674
Db	3942	---TATCCGTCCGCTCCCTCGTAAAGCTCGGCGCTGATCCGGTCCGCGTGGCCCGACG	3886
QY	1675	AspThralAlaIleuGluAsnArgIlnThrIleIleAsnAspTYrIleThrSerGlnI	1694
Db	3885	AACACTCCGGCGTGGCGCCCGCGAGTGAATCGGGGAGCACTCGGCGCTCTCCAGGGCC	3826
QY	1695	MetHisIAsnThralThr-----AlaMetAlGlnArgAla	1707
Db	3825	TCCCAACGACGCTTCAGACGACGAGCGCTCGCGGCTCAATGGCCGTGGCTCGCGGCG	3766
QY	1708	AspMetLeuArgIyLeuSerProArgGluSerLeuAlaIleAsnTYrAlaIalGly	1727
Db	3765	GATATTC-----CCGAAAGACTCC-----GCGTGAATCGGGCGCC---	3720
QY	1728	ProArgIyIleIleAspLeuSerGlnValProHisIleuProValLeuValProProThr	1747
Db	3729	-----GAGTCGAGAAACCGCGCTCCCGCACATAGTCTTCGGCGGCGTC	3685
QY	1748	ProGlyI-----ThrProAlaThralMetAspArgLeuAla-----Tyx	1760
Db	3684	CCGGGCTTCGGGCTGTCAGACCGCTTCAGAGTCCACAGCCCGGTGGTGGGAAAGAGAG	3625
QY	1761	LeuProThralaProGlnProPheSerSerArgHisIserSerSerProLeuSerProGly	1780
Db	3624	ATGCGCTCGGCGCGGACCTCCAGCAGAGCTCCAGAGAGCGCGCGCGCATCGACGCCCG	3565
QY	1781	GlyProThrHisIleuThrIlyProThrThrHiserSerArgIyValuArgAspArg	1800
Db	3564	GGAAGCGGACCGGACCA-----CCACAGATGGCGATCGGGTGTGTGTGTCCGGCTCG	3508
QY	1801	AspArgIyIuArgAspArgAspArgIyValuArgIyIuSerIleLeuThrSerThThrThx	1820
Db	3507	CCCCGGCGCTCGACACCCCTTACCGGCGGCGAGCAGAGACATCCGGGGCCACCGGCAAC	3448
QY	1821	ValGluIhIalPro-----IleThrArgProGlyThrGlnGlnIserSerGly	1836
Db	3447	GAGTCTTCGGCGCGCGCGCTTCCACGATGTGCTCGCGCAGGACCGCGGGGTGGC	3388
QY	1837	SerSerGlyIuSerSerGlyIyGlyIyGlyIySerSerSerArgProIaIserHisSerHis	1856
Db	3387	AGCTCGTAGACACAGCGTGGCGGCGGAGGGGAGCCCGGAGGCGCTTACGAGAGTCCGAC	3328
QY	1857	AlaHisGlnHisSerProIleSerPro-----ArgThrG1	1868
Db	3327	AGCTCGACCGCGGTACCG---AGTCAATGCGCTGGAGAGTAAAGTGCAGTCCGAGCG	3271

QY	1668	nAeRAlaLeuGlnGlnAArgProSerValLeuHisAenThrClyMetuVgLyLeIleTh	1868
Db	3270	ACCTCGCGGGAGTGGT	3252
QY	1888	rAlaValGluProSerLysProThValLeuArgSer	1902
Db	3251	-----GCGCCAGACCGCGCGCGTCTCCGGACACAGCGGACCGACTCTCGGGG	3199
QY	1902	rThSerSerProValAArgProAlaValThrPheProAlaThrHisCysProLeuG1	1922
Db	3198	GACATCGGCTCCGGTACCGCCGAGT-----CCGGCTCGTGGTCCGTCTCCCTGGG	3148
QY	1922	yG1YThrLeuAerGlyValYrProThrLeuMetGluProValLeuLeuProLysGluAl	1942
Db	3147	GAGCTCC-----CTCGGTCTC	3130
QY	1942	aProArgValAlaArgProGluAArgProArgAlaAerThrGlyHisAlaPheLeuAlaLy	1962
Db	3129	TCCGGGTCGGGGAGACGCTGTGCCCCACG-----CATGGTGGCGGTACA	3082
QY	1962	sProProAlaArg-----SerGlyLeu-----	1965
Db	3081	CCCCCGCCGGGTGGCGCGGTACCGCGCGCGTCAAGCGCGGCGACAGTGGTAAAGCGGT	3022
QY	1969	-----	1965
Db	3021	TGGAAGGGGTAGTGGAGGTGATGCGCGCGCGGTCTCCGGTGTGGATGCGGTG	2962
QY	1970	-GluProAlaSerSerProSerLysGlySerGluProArgProLeuValProProValSe	1989
Db	2961	GCGTCCAGCGCGCGCGTCCGGGTGACACGCTCGCACGCGCTCGACACCTCCCGC	2902
QY	1989	rGlyHisAlaThrIleAlaArgThrProAlaLysValAenLeuAlaProHisAlaSerPr	2009
Db	2901	TCCCGACGGCGCC-----GCCGATGGCGCGGACCA-----CCACCATATCATACCGGC	2851
QY	2009	o-----AerProProAlaProProAlaSerAlaSerAp-----	2020
Db	2850	CCCAAGCACTACACCGCGCATCCCGGTACGACCCCATAGCGGACCACTCCACCAAGCTC	2791
QY	2021	-ProHisArgGluLysThrGlnSerLysProPheSerIleGlnIleuGluLeuAArgSe	2040
Db	2790	CCCAACCCAGAACACGACGACACCCACCCCA-----TCCGCGAAAGCACCGCTCCCGC	2732
QY	2040	r-----LeuGlyTrpHisGlySerSerTySerProGlyGlyAlaGluProValSerPr	2058
Db	2731	ATAGCGACACCAATATCCCGGATCCCACTCCCTCC-----GACCCACACACCC	2681
QY	2058	oValSerSerProSerLeuThrHisAerLysGlyLeuProLysHisLeuGlnIleuA	2078
Db	2680	ACCGGACACACCGGACACCAACACCAACCGGCG-----CACCGACCGAACTCA	2628
QY	2078	pLysSerHisLeuGlnGlyLeuLeuAArgProLysGlnProGlyProValLysLeuGlyG1	2098
Db	2627	ACGACTCCAC-----AACCCCGCGCAACTCCCA	2598
QY	2098	yGluAlaAlaHisLeuProHisLeuAArgProLeuProGluSerGlnProSerSerSerPr	2118
Db	2597	ACACCGGCTCATACACCGGAAGAAACCATCGACATCCACACCG-----CC	2547
QY	2118	oLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValAlaThrLeuAlaGlnI	2138
Db	2546	GGACTCCACCCCATCCCTCCCATCGAGGAGAACACCACTC-----CAGCA	2499
QY	2138	sLieserGluValIleThrGlnAerTyrThrArgHisHisPro-----GlnGlnLeuSe	2156
Db	2498	CACCCAC-----ATCACCGGACAAACCAACCGACCGCGCGCATTCACCGCGGCAACG	2445
QY	2156	rAlaProLeuProAlaProLeuTySerPheProGlyAlaSerCysArgProValLeuAerLe	2176
Db	2444	ACACCACTCCCTCCCAACCTCCCAACACCCCGGACACCGA-----CT	2400
QY	2176	uArgArgProProSerAerLeuTyrLeuPro-----ProProAerHis-----	2190

```
Db 2399 CCGAGCGCC-----CACCGACACATCCCTCCACCCACCGGCAACCCACCA 2352
Qy 2191 ----GlyAlaProAlaArgGlySerProHisSerGluGlyGlyLeuArgSerProGluPr 2209
Db 2351 TCACCCACCGCGCGCGCACCAACCGACCGCATCCACCAACCAACACCGCGCA 2292
Qy 2209 oAnlysrThrSerValLeuGlyGlyGlyLeuArgGlyLeuProValSerProGlu 2229
Db 2291 CATACGACGACGACCTC-----CCCCA 2268
Qy 2229 uGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeu----- 2245
Db 2267 CCGAATGACCA-----CACACGACACCTCCACACCTCGAGCTCCAAAGCCCGGA 2214
Qy 2245 ----- 2245
Db 2213 ACACGCGCACTTCACCAACGCAACAGCAGCGCTGAGCAAACTCCGCGACCCACCAACC 2154
Qy 2246 ---TyrArgAspGlyGluGlnThrGluProSer---ArgMetGlySerLys-Ser--- 2261
Db 2153 CACCCACAGAACCGGAACCGAACCGACCCGACACCGGACCGGACCGCATCAGCCC 2094
Qy 2262 ----ProGlyAsnThrSerGlnProProAlaPheSerLysLeuThrGluSerAsnS 2280
Db 2093 CCGCACCCCGCTCCACGTCGACGTCGCG-----AACATCACATCCCGCAACG 2046
Qy 2280 eAlaMetValLysSerLysGlyGlnLilLeaLysLysLeuAsnThrHisAsnArgA 2300
Db 2045 ACCGACATCCACCTCCCGACATGACACAAACCTCATCCAGCACTCCGCAAAACCC 1986
Qy 2300 snGluProGluTyrAsnLysSerGlnProGlyThrGlu-----IlePheA 2315
Db 1985 CACCCCGCATATGACACACCGCCCATACCAACCTATGCGACCGCAGCCGTGACAA 1926
Qy 2315 snMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluH 2335
Db 1925 ACACCCCGACCCGACCGGACAGCC----- 1900
Qy 2335 iAlaSerThrAsnMetGlyLeuGlnAlaIleIleArgLysAlaLeuMetGlyLysTyrA 2355
Db 1900 ----- 1900
Qy 2355 spGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerA 2375
Db 1899 ---CGCCCGACCGCACACCCCGACACCTCCCGGACGAAACCCACCAAGACGACGA 1842
Qy 2375 lAsnLeuPro----- 2378
Db 1841 CGAGCTCCCGACCGCACACGACGACCGCAGCTGCTGAAAAGATCCGCGTCG 1782
Qy 2379 --AlaAlaMetProIleThrAlaIleAspGlyArgSerAspHisThrLeuThrSerProG 2398
Db 1781 AACCGACGCAATTAAGACATATCCCGCGGATCCGCGCCACCCCTCCACATATCAACCA 1722
Qy 2398 lYcGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProA 2418
Db 1721 ACCGACCGCGCTCGCCCG-----AACGCGCCCTCCGACCGC---GGGACAAACAA 1671
Qy 2418 lAProGlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluGlyA 2438
Db 1670 GCGCGCGGACGTCGGAAGAGAGCGGCGACGACGACGCGGCGGCGGAGAGGCA 1611
Qy 2438 spCyAsnArgArgTrn-----ProLeuThrAsnArgValTrpGluAspA 2453
Db 1610 GGTTGTTCCAGACACATATGCACTTGTCACCGATCCGCAAC-----GAACTGA 1560
Qy 2453 rGProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnA 2473
Db 1559 CACCCCGACGAGAGGGGCGCGCC----- 1534
Qy 2473 lAglyValMetAlaSerProProProProGlyLeuPro---AlaGlySerGlyProLeu 2491
```

```
Db 1533 --GGGCGGACGGCGACGCTCGCGGACGCGTCTGCACCTTCACGCGGAGCTGTGCAGCG 1476
Qy 2492 AlaglyProHisAlaIleTrp 2498
Db 1475 GGATGCGCGGAGAGGGGGTGG 1455

RESULT 12
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 8,85e-19 Length: 4403765
Score: 668.50 Matches: 617
Percent Similarity: 32.38% Conservative: 255
Best Local Similarity: 22.91% Mismatches: 1044
Query Match: 5.06% Indels: 793
DB: Gaps: 113

US-09-522-753-5 (1-2517) x US-09-103-840A-2 (1-4403765)
Qy 4 SerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrProProHisSer 23
Db 2794906 TCGACGAGAGCCCGCATCGCTGCTGCGCGCCGACCGCAATACGCGATCCACCGGTCAGC 2794965
Qy 24 -----LeuSerTyrProValGlnIleAlaArgThrHisThrAsp 36
Db 2794966 AGCGGAAACCGGTCATGCAAGCTGTGACGATTTGCTGAGCGGCGGACGCGACGAC 2795022
Qy 37 ValGlyLeuLeuGluTyrGlnHisSerArgAspTyrAlaSerHisLeuSerProGly 56
Db 2795023 -----CGCGACGCTGCAAGTTGATCCGACCGGA 2795052
Qy 57 SerIleIleGlnProGlnArgArg-ArgProSerLeuLeuSerGlnPheGlnProGlyAs 76
Db 2795053 -----ATGCGCGTGAAGTGGTGGGACAGATCTCG-----TCACCAAGGGG 2795091
Qy 76 nGluArgSerGlnLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuG 96
Db 2795092 AGTTGTCGACATGATCTGATGTCGCGCGCGCCGCTCAGACGCGCGGTGTA----- 2795146
Qy 96 lYlsSerGluMetGluPheIleGluSerLysArgProArgLeuGluLysLeuProAspP 116
Db 2795147 -----CAATCGATCGGCTCGCGCTGCG 2795169
Qy 116 oLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGlnAspLeuTh 136
Db 2795170 CCAGCGGGGAAACCGCGCAATCTGCTCACCGGCAC----- 2795206
Qy 136 rLysAspArgSerLeuThrGlyLysLeuGluProValSerProProSerProHisThr 156
Db 2795207 -CGCATCGTTCCGCGCTGTGCG----- 2795230
```

QY 156 rAspProGluLeuGluLeuValProProArgLeuSerLeuGluLeuLeuGlnAme 176
 Db 2795231 -----CAATACCTCAGCGCTGGGCAAGCCCGAG----- 2795260
 QY 176 rAspArgValAspArgGluLeuMetValGluGlnGlnLeuSerLeuValbly 196
 Db 2795261 -CAACCGAGCATCAGACCGCGCACCGGTCAGCAAGTCCTCCAGTTGTCAGACTA 2795319
 QY 196 sGlnGlnGlnLeuGluGluGluValAlaLeuValProProGluProGluValblyAsp 216
 Db 2795320 CCAAG-----CATGGCCGGTCGCGA 2795340
 QY 216 oProProlLeuSerLeuValblyArgSerLeuValGlnLeuLeuValblyArgLeuValbly 236
 Db 2795341 TAGCCGCAACAAATGTCATCCAGCGCGCGCGGTATCCGGCAACCCCAAAACC 2795400
 QY 236 blybAlaGluAlaAlaHisArgLeuLeuGluGluValblyProGlnValGluLeuPro-- 255
 Db 2795401 GCGCGCGCGCATCGACACGCGCGGTGATCGCGCGCGCATGTCATACCAAA 2795460
 QY 256 -----LeuTyra 258
 Db 2795461 CCCGTCGGAATACCTCGGCAACGCGCTCGGACCTGTGTCAGCGCGGTCTTTC 2795520
 QY 258 ngInProSerAspThrArgGlnTyriArgLeuAlaLeuValblybAlaMetArgly 278
 Db 2795521 CGACCCCGCGACACCGGTAAAGGTGACCAACCTTTGAGAGTCCAGACGCCACGAC-- 2795578
 QY 278 blybLeuLeuLeuTyriPhelybArgArgbAlaAlaArgbGlnTyriTyriblyblybph 298
 Db 2795579 -----TTCCGCGCATCTTCCTCGACGCGCCACACCGCATGAGT----- 2795617
 QY 298 eCySGlnArgTyriAspGlnLeuMetGluAlaLeuGluValblyValGluArglyLeu 318
 Db 2795618 -----CTGGCGGAAAGACAGTGGCGCAACGA 2795646
 QY 318 naenProArgArgAlaArgAlaLeuSerLeuValArgGluTyriTyriGluValblyblybph 338
 Db 2795647 CTTTCGGGCGCGCACCGGCGGGAACCGCTTGTCAATCAGAGGTGACACAGCTGCACCA 2795706
 QY 338 ogLuiLeuArgblybGlnArgGluLeuGln-----GluArgMetGlnSerArgva 354
 Db 2795707 CCCGTTCCGGTCGGGACAGGTCTCCAGCGGTAGGTACGAGGTCTGTCAGCCACGCGT 2795766
 QY 354 l---GlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGlnHisGluValSe 373
 Db 2795767 CCTTGGGCGAG-----CAGGTGAG 2795784
 QY 373 rGluLeuLeuAspGlyLeuSerGluGlnGlnLeuGlnLeuGlnLeuMetArgGlnLeuAl 393
 Db 2795785 CAACGAGATCGCTGTAGTTCCCGACACACGAGTCTG---GCCCGCGTGGCCAGCTCGC 2795841
 QY 393 aValIleProProMetLeuTyriAspAlaAspGlnGlnArglyLeuValblybHele----- 410
 Db 2795842 GCGAG---CCGGGCGGTGCGGTCTCATGTGCGGCTTACGACAGTTGCTCTGCGGTGACG 2795898
 QY 411 -AsnMetArgGlyLeuMetAlaAsp-----ProMetly 421
 Db 2795899 ACACCTTCCCGGTGTGATGTCATGCGACGCGGATCGGTGCGACGCGCGCGCTGCA 2795958
 QY 421 sValTyriLeuAspArgGlnValblyMetAsnMetTyriPserGluGlnGluValblyblybph 441
 Db 2795959 AGCCCAAGGGCGGACGCAACGCGCTCGATCCGGGCGGAAAGCGCAACCAAGAGCTGTCCG 2796018
 QY 441 gGluValblybMetGlnHisProLybAsnPhelyLeuLeuAlaSerPhelyblyblybly 461
 Db 2796019 CTTGCGCTGTTCAGACCGGCAAAACCCGCGGTGTCGCAACCAATTCCGTGAGGTTTC 2796078
 QY 461 sThrValAlaGluCyValLeuTyriTyriTyriLeuThyLeuValblyblyblyblyblybse 481
 Db 2796079 G----- 2796079

QY 481 rLeuValArgSerTyriArgArgArgGlyLySerSerGlnGlnGlnGlnGln 501
 Db 2796080 -----GTCCAGTTGGCGATGCGCGTGTCTGTCAGCTAGACCCCGGAGAGGTGTCCGC 2796135
 QY 501 ngGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 521
 Db 2796136 CTTGATATCGGCGACAGCAACGTCA-CCGTGCCG----- 2796171
 QY 521 pGluValblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyb 541
 Db 2796172 -----TCGGTACAGCTGCTTACACCATTCGCGCTCCAGTTCACAGGTACCACT-- 2796222
 QY 541 pLySGluAspLeuLeuValblyblyblyblyblyblyblyblyblyblyblyblyblyb 561
 Db 2796222 ----- 2796222
 QY 561 sGluAlaValAlaSerLeuValblyblyblyblyblyblyblyblyblyblyblyblyblyb 581
 Db 2796223 -----CGACGCGCGGTGA 2796236
 QY 581 gLileThrArgSerMetAlaAsnGlnAlaAsnSerGluGluAlaAlaLeuThyrProGlnInse 601
 Db 2796237 -ATCTTGCTCATGCTAGCC-----AGCATGAGCCAGCGCTAG 2796274
 QY 601 rAlaGluLeuAlaSerMetGluLeuAsnGlnSerSerArgTyriThyrGluGluGluMet-- 620
 Db 2796275 CGCATTAACATCGGACACTGCGCTTGAATGCTGGAATCTCTTGCGCGCGGTGTCATGTC 2796334
 QY 621 -----GluThrAlaValblyblyblyblyblyblyblyblyblyblyblyblyblyb 635
 Db 2796335 TTCTACGCGCTTATGTCATGTCATTCGCGCACCGCGGTGTCATTCGCGCGCGG-ACCTACAG 2796393
 QY 635 rAlaIleAlaArgMetValGlySerTyriThyr-----ValSerGlnCyValblyblyblyblyb 654
 Db 2796394 T-----GCCGCTGCTCGCTGCGCAATTGATGCCCACTG----- 2796433
 QY 654 eAsnTyriLyblybArgGlnAsnLeuAspGluLeuGlnGlnHisblyblyblyblyblyblyb 674
 Db 2796434 -----TGCCACACATCGCCCGCAAGAAC 2796456
 QY 674 ulblyb 694
 Db 2796457 GACACGACCGCATCGTGGGCGGTG-CCGTGCAAGTGGCGCGGTGACCGATCGCGGAGTTC 2796515
 QY 694 a-----AlaPhyProProValValGluAspGluGluMetGluAlaSerGlyva 710
 Db 2796516 ACCGCGCGGATGGGTTTCGAAGGCTAGGAAATGATGAGATGGGCGCGCTCGGC 2796575
 QY 710 l-----SerGlyAsnGluGluGluMetValGluGluAlaGlu 722
 Db 2796576 CGCTTCGCGCTAACCCCAACCATCGGAAACGAGCTTACAGTTCGTGACCATGAGTGG 2796635
 QY 722 uAlaLeuHisAlaSerGlyb---AsnGluValProArg-----GlyGlu 725
 Db 2796636 AGCGACGACATCGCGGTGGCGGCAACGAGGAGCATCGGATCTCTGACCATGAGTGG 2796695
 QY 735 uCySerGlyProAlaThyrValAsnAsnSerSerAspThyrGlnSerIleProSerProH 755
 Db 2796696 GTGTTCTGCTACTAGCTTA-----ACGGAGATGAGAGGCGCGCTTCGCGC-- 2796744
 QY 755 sThrGluAlaAlaValblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyb 775
 Db 2796745 -----CTTCAACGCGC----- 2796756
 QY 775 aAspGlyProProProGlyProProProThyrProProArgArgThyrSerArgAlaProIleGlu 795
 Db 2796757 -GAACCGCGGAGCGCGGTGCGCTCTGCGCGGACCAACCGACTGCGCGCACGA 2796815
 QY 795 uProThyrProAlaSerGluAlaThyrGlyAlaProThyrPro-----ProProAlaProP 813
 Db 2796816 GCGAGCGCGCGGAGGTCAAGCACCGGACCGGCTCGGCTCACTTCGCGCGCGC 2796875
 QY 813 oSerProSerAlaProProProValValProLyblyblyblyblyblyblyblyblyblyblyblyblyblyb 833

```

Db 2796876 GTCCTCCGCGTACCGCGCCGATGCTGG -AGCGGAGGGGCGCGGAGCCGCGGAGC 2796934
      ||||| ..... ||||| ..... |||||
Qy 833 AlaProPro-----ValGluGluGluGluGluGluProProAlaAla 849
      |||||
Db 2796935 CCGCGGTCCGCGCGCTCCGCTCCACCAACCGGGGCGCCACCGTCTCCGCGCGCCAC 2796994
      |||||
Qy 849 uGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGlu 869
      |||||
Db 2796995 CCAAGCGCGCGTT--GCGCGCGTCCACCGCGCGCGCGCTCCGCTGCATCAGCAAGC 2797051
      |||||
Qy 869 uAlaGluGluGluProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGlu 889
      |||||
Db 2797052 TCGCGTTCGCGTCCCGACCGTACCGCGCTCCGCGCTCCGCTCCGCTACCGCGGGG 2797111
      |||||
Qy 889 yAlaLeuLysAlaGluLysGluGluGluGluGluGluGluGluGluGluGluGluGlu 905
      |||||
Db 2797112 CGCC-----GAAAGCGCGCGCTTCGCGTCCGCGCGCTTCGCGCGCGCGCGCGCGCG 2797156
      |||||
Qy 905 rAlaLysSerSerGlyAlaProGluAspSerAspSerSerAlaThrCysSerAlaAspG 925
      |||||
Db 2797157 CCGCTTGTTCGCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2797216
      |||||
Qy 925 uValAspGluAlaGluGlyLysLysAspLysAspLysSer- ProAspProSerLeu 945
      |||||
Db 2797217 GCTC--TCCGTCAAATCCGCGAGATAGGTAAAGCGGTACACCGCGCGCGCGCGCGCGCG 2797273
      |||||
Qy 945 euthrProthrGlyAspProAspAlaAspAlaSerProGluLysProLeuAspLeuLysG 965
      |||||
Db 2797274 CAGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2797324
      |||||
Qy 965 LheuLysGlnArgAlaAlaAlaProProIleGlnValThrLysValHisGluProP 985
      |||||
Db 2797325 TCGCGACCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2797363
      |||||
Qy 985 rGArgGluAspAlaAlaProThrLysProAlaProProAlaProProProGluAsn- 1004
      |||||
Db 2797364 CGTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2797423
      |||||
Qy 1005 -----LeuGlnProGluSerAspAlaP 1012
      |||||
Db 2797424 GCGCGCTGCTTCGTTAAATGTCAAAGCGCGTCAACCGCGGTACCGCGCGCGCGCGCGCG 2797483
      |||||
Qy 1012 rGlnGlnProGluSerSerProArgGlyLysSerArg-----SerProAlaPro 1029
      |||||
Db 2797484 CCGCTAGGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2797543
      |||||
Qy 1029 rGAla-----AspLysGluAlaAspAlaP 1036
      |||||
Db 2797544 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2797603
      |||||
Qy 1036 LAlaGluAlaGlnLysLeuProGlyAspPro-----ProCysThrThrSerG 1052
      |||||
Db 2797604 GGTGGAGTGGCGGACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2797659
      |||||
Qy 1052 LysLeuProPheProValProArgGluValLysLysAlaSerProHisAlaProAsp 1072
      |||||
Db 2797660 -----CGGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2797707
      |||||
Qy 1072 rGAlaPheSerLysAlaProProGlyLysAlaProLeuProLeuGlyLysHisAspThr 1092
      |||||
Db 2797708 -----TGCACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2797740
      |||||
Qy 1092 LArgProValLeuProArgProProThrLysSerAspProProProLeuLysSerSer 1111
      |||||
Db 2797741 -----CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2797783
      |||||
Qy 1112 AlAlaHisProSerValLeuGluArgGlnLysLysAlaLysSerGlnLysMetSerVal 1131
      |||||
Db 2797784 CCGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2797813
      |||||
Qy 1132 GlnLeuHisValProTySerSerGluHisAlaLysAlaProValGlyProValThrMetGly 1151
      |||||

```

```

Db 2797814 -----CACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2797866
      |||||
Qy 1152 LeuPro-----LeuProMetAspProLysLysValAlaProPheSerGlyValLysGlnGlu 1170
      |||||
Db 2797867 CCTCCGAGGGGACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2797911
      |||||
Qy 1171 GlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGln 1190
      |||||
Db 2797912 CCGCGTCAACCGATGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2797959
      |||||
Qy 1191 GluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyLysSerIleThrLys 1210
      |||||
Db 2797960 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2798018
      |||||
Qy 1211 GlyIleProSerThrArgValProSerAspSerAlaIleThrTyArgGlySerIleThr 1230
      |||||
Db 2798019 GGTCTCGCGCGTCCCGGATAGCGCGCGCGTGGCGCGGTACCGCGCGCGCGCGCGCGCG 2798078
      |||||
Qy 1231 HisGlyThrProAlaAspValLeuTyLysGlyThrIleThrArgGlyLysGluAsp 1250
      |||||
Db 2798079 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2798108
      |||||
Qy 1251 SerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTy 1270
      |||||
Db 2798109 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2798165
      |||||
Qy 1271 GluGlyLysLysGlyHisValLeuSerTyGluGlyGlyMetSerValThrGlyCysSer 1290
      |||||
Db 2798166 -----GCGCGCGAGGATCGCGTCCGCGTACCGCGCGCGCGCGCGCGCGCGCG 2798222
      |||||
Qy 1291 LysGluAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLysAlaGln 1310
      |||||
Db 2798223 -----GGGTCACCGCGCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2798249
      |||||
Qy 1311 TyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeu 1330
      |||||
Db 2798250 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2798306
      |||||
Qy 1331 MetGlyArgAlaIleProProGluArgHisSerProHisLysLeuLysGluGlnHisLys 1350
      |||||
Db 2798307 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2798362
      |||||
Qy 1351 IleArgGlySerIleThrGlnGlyIleProArgSerTyValGluAlaGlnGluAspTy 1370
      |||||
Db 2798363 -----CCGTCCCGAGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2798395
      |||||
Qy 1371 LeuArgArgGluAlaLysLeuLysLysLysLysLysLysLysLysLysLysLysLys 1390
      |||||
Db 2798396 CGTCCCGCGCGATGTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2798454
      |||||
Qy 1391 ArgAspLeuThrGluAlaTyLysThrGlnAlaLeuGlyProLeuLysLeuLysProAla 1410
      |||||
Db 2798455 CACACGGTGGCGCGCGATGTCGATGATA--CGTGGCGCGCGCGCGCGCGCGCGCG 2798496
      |||||
Qy 1411 HisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArg 1430
      |||||
Db 2798497 -----GCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2798505
      |||||
Qy 1431 GlnGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGly 1449
      |||||
Db 2798506 CACCGCTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2798561
      |||||
Qy 1450 SerIleThrGlnGlyThrProLeuLysTyAspThrGlyAlaSerThrThrGlySerLys 1469
      |||||
Db 2798562 GCGGTACCGCGACACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2798594
      |||||
Qy 1470 LysHisAspValArgSerLeuLysGlySerProGlyArgThr-PheProProValHisAsp 1489
      |||||
Db 2798595 -----GGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2798630
      |||||
Qy 1489 oLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyGluGluSerLeuLys 1509
      |||||
Db 2798631 C----- 2798631
      |||||

```



```
Db 2800312 CCGCCGTTCCCGCCCTGACCGGCGCGTGGCCGCGCGGACGACATTTG-GCCCGCGC 2800370
Qy 2194 aArgGlySerProHisSerGluGlyGlyAspArgSerProGluProHisLeuThrSerA 2214
Db 2800371 GCGACACACCGGCGCGCGCCCGCGCGCTCCGCGACCGCCACCGAG----- 2800422
Qy 2214 LeuGlyGlyGlyGluAspGly11eGluProValSerProProGluGlyMetThrGluPr 2234
Db 2800423 -----CCCGGCTGCGCACCGTACCGCGCGCGCGCGC 2800454
Qy 2234 oGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluPr 2254
Db 2800455 GTACACCGCGCGCGCGCACTGCACCGCTTG-----CCCGCGGACCGCGC 2800499
Qy 2254 oSerArgMetGlySerLeuSerProGlyAsnThrSerGlnProProAlaPhePheSerIy 2274
Db 2800500 GTACACCGAGCGCTCCACCGACCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 2800559
Qy 2274 SteuthrGluSerAsnSerAla-MetValIySerIySylGlnGlu1LeAsnIyVal 2294
Db 2800560 GACATATCCCATATGTCACCGGACCGGACCGCGCGCGCGCGCGCGCGCGCGCTT 2800619
Qy 2294 euAsnThrHis-----AsnArgAsnGluProGluTyrAsnI 2306
Db 2800620 TCCGCGCGGACCGCATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGATACCGG- 2800678
Qy 2306 IeSerGlnProGlyThrGluIlePhe-AsnMetProAlaIleThrGluIyLeuMet 2325
Db 2800679 -----CCGCGCGCATTTATTCGGTAGCCCGCGGACCGCGCGCGCGCGCGCGC 2800727
Qy 2326 ThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIle 2345
Db 2800728 ACCGGACACACCGGCGCGCGCGCGCGGACCGGACCGCGCGCGCGCGCTTACC----- 2800781
Qy 2346 I1eArgIyAlaLeuMetGlyIySylTyrAspGlnTPrGluGluSerProProLeuSerAla 2365
Db 2800782 -----GAGAAAGCGACCTTCCCGCGGACCGC 2800805
Qy 2366 AsnAlaPheAsnProLeu---AsnAlaSerAlaSerLeuProAlaIleMetProIleThr 2384
Db 2800806 GTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 2800865
Qy 2385 AlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyIyValIyAlaIyS 2404
Db 2800866 GCGCGC-----ACGTCGTCGACACCGGCAACACAGCGCGCGCGCGCATC 2800907
Qy 2405 ValSerGlyArgProSerSerArgIyValIySerProAlaProGlyLeuAlaSerGly 2424
Db 2800908 CTGACCGGCGCGGTAGCACCGCTTGCGCGCGCGCACTACACACCGCGCGCGCGCGCGC 2800967
Qy 2425 -----AspArgProProSerValSerSerValHisSerGluIyAspIySylAsnArg 2442
Db 2800968 CCCATGACCAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 2801024
Qy 2443 ThrProLeuThrAsnArgVal-TPrGluAspArgProSerSerAlaGlySerThrProPh 2462
Db 2801025 ACCGGACACACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 2801084
Qy 2462 ePro-----TyrAsnProLeu11 2468
Db 2801085 GCCACACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 2801144
Qy 2468 eMetArgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeuProAlaGlySe 2488
Db 2801145 CCGCGCGGATTCCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 2801204
Qy 2488 rGlyProLeuAlaGlyPro 2494
Db 2801205 TAAACCGCGCGCGCGCGCG 2801223

RESULT 13
US-09-103-840A-1
```

```
/ Sequence 1, Application US/09103840A
/ Patent No. 6294328
/ GENERAL INFORMATION:
/ APPLICANT: FLEISCHMAN, Robert D.
/ APPLICANT: WHITE, Owen R.
/ APPLICANT: FRASER, Claire M.
/ APPLICANT: VENTER, John C.
/ TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
/ FILE REFERENCE: 24366-20007.00
/ CURRENT FILING DATE: 1998-06-24
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 1
/ LENGTH: 4411529
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ OTHER INFORMATION: H37Rv
/ US-09-103-840A-1
```

```
Alignment Scores:
Pred. No.: 1,286-18 Length: 4411529
Score: 665.00 Matches: 627
Percent Similarity: 33.04% Conservative: 236
Best Local Similarity: 22.44% Mismatches: 1111
Query Match: 5.03% Indels: 776
DB: 3 Gaps: 122

US-09-522-753-5 (1-2517) x US-09-103-840A-1 (1-4411529)
Qy 20 PropionHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeu 39
Db 2798750 CCGCGGACGCGGATGACCGC-CGCGCGCGACGAGCGCG-----ATC 2798790
Qy 40 LeuGluTyrGlnHisHisSerArg-----AspTyr 49
Db 2798791 CTCGATGTCGCGCGCTATGACGAGCGCTTCTGGCATCATGTAGCGGTCGGTAGTGC 2798850
Qy 50 AlaSerHisLeuSerProGlySerIle-----IleGlnProGlnArg 64
Db 2798851 GGCACGCGCATTCATGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 2798909
Qy 65 ArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnIleuHisIleu 84
Db 2798910 CCGCGCGGATGTGATGTGATGTGCGCGGCGCGGCGGACGCGCGCGCGCGCGCTT 2798969
Qy 85 -----ArgProGluSerHisSerTyrLeuProGluLeuGlyIySer 98
Db 2798970 GACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 2799029
Qy 99 GluMetGluPheIleGluSer-----LysArgPro-----ArgLeuGluLeu 112
Db 2799030 CCGTATCGCATTCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 2799086
Qy 113 LeuProAspProLeuLeuArgProSerProLeuLeuAlaThrGlyIleProAlaGlySer 132
Db 2799087 -----CGTCGCGCTGTGTGACCGCGCGCTTCATGCGCGCGCGCGCGCGCGCGC 2799131
Qy 133 GluAspLeuThrIyAspArgSerLeuThrGly-----Lys 144
Db 2799132 CGTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 2799191
Qy 145 LeuGluProValSerProProSerProProHisThrAsp---ProGluLeuGluLeuVal 163
Db 2799192 CATCGCGGATTCACGACACTTCTCCAAAGCATCTGCGCGCGCGCGCGCGCGCGC 2799251
Qy 164 ProPro-----ArgLeuSerIySylGluIleuIleGlnAsnMetAspArg 178
Db 2799252 AGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 2799311
Qy 179 ValAspArgGluIleThrMetValGluGlnGlnIleSerIyLeuLeuIySylSylGlnIn 198
DB: 3
```

D	2799312	GCAGATTCACCAACCTCGTAGGGCTGAACGTACACCGCGCGGCTGCGCTTGGGACCGT	2799371
O	199	GlnLeuGluGluGluAlaAlaAlaLeuProGluProGluIlyProValSerProPro	218
D	2799372	CGAGATCAAAACCCGCTCGGAAACCCGACATCGCGCA-----ACAGACCTGTCTCG	2799425
O	219	IleGluSerIlyValSerSerLeuValGlnIleIle-TyrArgGluAsnArgIlyVal	238
D	2799426	GTC-----CGGTACGACGACCGCATGTGACCACTGCACGGAAAGCCCGCATCTCT	2799473
O	238	AGLValAlaIleValArgIleLeuGluGluIlyLeuGluProGlnValGluLeuProLeuTyrAs	258
D	2799474	GCTGGCGGCGGACCGCAATACCGATCCACCGGTACGACGCGGAA--CCGGTCATGCA	2799530
O	258	nglnProSerapThrArgGlnTyrHisGluAsnIleIysIleAsnGlnAlaMetArgIly	278
D	2799531	AGCTGTGACGATTTTGGTCA-----	2799552
O	278	blYleuIleLeuTyrIleValSerArgAsnHisAlaArgIlyGlnTyrIlyGlnIlySph	298
D	2799553	-----CGCCAGGCGACGACCGCGGACG-----	2799576
O	298	eCyagGlnArgTyrIleArgGlnLeuMetGluAlaLeuGlnIlySylValGlu-----	314
D	2799577	-TGC-----AAGTTCATCGCCACGCGGAATGCTGTCGATCGGTGACAGATCTCGGTACCA	2799632
O	315	-----ArgIleGluAsnAsnProArg-----	321
D	2799633	GGGGAGGTTTCGGGACATATCTCGATTTGGGCGGCTTCACGACCGGCTCGTGA	2799692
O	322	-----	2799693
D	2799693	ACAATCGATCGCTCGCGCGCGCCAGCGGGGAAACCGCCAAATCTGCTCACCGGCA	2799752
O	333	rgIuIlySglIlePheProGluIleArgIlySglIleArgIlyGlnIlyArgMetGlnSerAr	353
D	2799753	CCGGATCGGTGTCCTCGGTGTTCCTCAATACCTCACCGCTGGGACAGCCCGGACAGC	2799812
O	353	gValGlyGlnArgIlySerGlyLeuSerMetSerAlaAlaArgSerGluHis--GluVa	372
D	2799813	CGAGATCAAGACCGCGCACCCCTCGACGAAGTCTCGCACTTCCAGCATACCGCA	2799872
O	372	IserGluIleIleAspGlyLeuSerGlnGlnAsnLeuGluIlySglIleMetArgIlyLe	392
D	2799873	TGCGCCGGTGCAGAT-----ACGCGGCAAAATGTGTCCACCGCTGAGAGGCGCGCT	2799926
O	392	uAlaValIlePro-----ProMetLeuTyrIlePheAlaAspGlnIleArgIle-----	407
D	2799927	GATCCGGCAACCCCAAAACCGCGCCGCGGATCGGACACGCGCGGTGTGATCG	2799986
O	408	-----LysPheIle	410
D	2799987	GCGCCAGTTGACTTACCAAAACCCGTCGGATTAACGTGGGCAACGCGCTGCGACCT	2800046
O	410	eAsnMetAsnGlyLeuMetAlaAspProMetIlyValTyrIleAspArgIleValMetAs	430
D	2800047	GTCGCGCAGGCGGTCTTTCGACCCCGCGCAACCGGTAAGGTACACCAACGTTTGA	2800106
O	430	nMetTyrSerGlnGlnIlySglIlyThrPheArgGlnIlySphMetGlnHisProIlyAs	450
D	2800107	CGTC-----CAGACGCCACACGACTTGTGACACCGCGCCACCA	2800154
O	450	npheGlyLeuIleAlaSerPheLeuGluArgIlyThrVal-----	463
D	2800155	G-----CCGAGTGAAGCTGGGCGCGAAGACAGTGGCAACAAACATTTCCGGG	2800202
O	464	-----AlaGluCysValLeuTyrTyrTyrIleThrIlySylAsnGlnIlySyl	480
D	2800203	TCCGACAGCGGCGGAACGCGTGTGACAGATCAGGGTGAACAAGTGCACACCAACCGTTCCG	2800262
O	480	ser-----LeuValArgArgSerTyrArgArgArgIlySylSerG	494
D	2800263	GTCGGGACAGTCTTCAGCGCGGTAGTACCGAAGTCTTTCAGGACAGCTCTTGGGCA	2800322

[illegible]

[illegible][illegible]

QY	54	serProgluSerIleIleGlnProGlnArgArgProSerIleuSerGluPheGln	73
Db	12977	AGCCCGCCGCCGCTCAGACAGCAGATGCGCACACCGGCTCGCGCCAGATG	12918
QY	74	ProGly-----AenGluArgSerGlnGlu	81
Db	12917	CCGGGCGCAGACGCCGCCAGCGTGGCGGTCGCCCGCTCACACAGACCGTGC	12856
QY	82	LeuHISleuAArg-ProGluSerHISerTYrLeuProGluLeuGlyUlysSerGluMetGln	101
Db	12857	GTCGAATGCGCGACCCCGCTGCTCTCCGGGATGCCGGCCGAGAGAGACCGGGGTCCG	12798
QY	101	uPheIleGluSerIleuTYrProArgLeuGluLeuLeuProAspPro-----	116
Db	12797	GCGCGTCCCGGCCCGCAG-----GAGAGCTGGCCGGCCGCTCGAGACCGCGG	12747
QY	117	-LeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuTh	136
Db	12746	CAGCCGAGGGGGCGCGCGGTGCGGTGCGGTCCGAGATGCCGTGGATGTCAGAGGGCGAA	12687
QY	136	rLYe-AspArgSerLeuThrGlyUlysLeuGluProValSerProProSerProProHISr	156
Db	12686	GCGGCGCGGGGTCTCGGACCGCGGAGCTCACCA-----GGCCCCACACCCCGCACA	12633
QY	156	hYrAspProGluLeuGluLeuValProProArgLeuSerIleuGluLeuIleGlnAsnM	176
Db	12632	CGCCAGGTCCCCCATCTCTCCCGGTACG-----	12601
QY	176	eYrAspArgValAspArgGluIleThrMetValGluGlnIleSerIleuLYeLYeLYeL	196
Db	12600	-----CGCGCGAC	12594
QY	196	yGlnGlnGlnLeuGluGluGluAlaAlaLysProProGluProGluLYeProValSerP	216
Db	12593	GGCTCCCCCTGGTAGCACTATGAGCGGTGTGTCTTCCCACTCTCGGCGCTCCAGCCAGTC	12534
QY	216	roProProIle-gluserLYeHISArgSerLeuValGlnIleIleTYrAspGluAsnArg	235
Db	12533	CTGGAGCGCGGTGAGACAGCGCTCCGTTGCTG-----	12501
QY	236	LYeLYeAlaGluAlaAlaHISArgIleLeuGluGlyLeuGlyProGlnValGluLeuPro	255
Db	12500	-----CCGCACCGACTCTCGGAGTGGCCCGCGC-----	12474
QY	256	LeuTYrAsnGlnProSerAspThrTYrArgGlnTYrHISgluAsnIleLYeIleAsnAla	275
Db	12473	-----CCCGAGAGACACCGCAGATC-----GACCGCTCCGGCGCGCGCGCGC	12429
QY	276	MetArgLYeLYeLeuIleLeuTYrPheLYeAspArgAsnHISAlaArgLYeGlnTrpLYe	295
Db	12428	CGCGCGCAGCGCTT-----CCGCACCGCGCGACTCTCGGGTACAGCGCGGT	12381
QY	296	GlnLYePheCYeGlnArgTYrAspGlnLeuMetGlnAlaLeuGlnLYeLYeValGluArg	315
Db	12380	GGGGCCACCGTCCGGCTGATCCCGCAGGCCCGCGCGTCCGGCGAGCGCGTCCGCTCGG	12321
QY	316	IleGluAsnAsnPro--ArgArgArgAlaLYeGluSerLYeValArg-----Glu	331
Db	12320	GCCGAGACCGCCAGAGCGCGCGGGCGCGGCGAGAGCGCGGCTGCACATCGAC	12261
QY	332	TYrTYrGluLYeAsnPheProGluIleArgLYeGlnArgGluGlnGlnIleuArgMetGln	351
Db	12260	GCGGTACAGGACCGTTCGCGCGCGCTCGGCGCGACAGACCTCCAGCTTCCGCGCGCC	12201
QY	352	serArgValGlyGlnArgLYeSerGlyLeuSerMetSerAlaIleArgSerGluIleGlu	371
Db	12200	CAGCGAGCAGATCACACGAGC-----CCCAAGAGAGGACCGCG	12162
QY	372	ValSerGluIleIleAspGlyLeuSerGluGlnGluAsnLeuGluLYeGlnMetArgGln	391
Db	12161	CGCGCGCAGCATGTTCGCGCGCTGAG-----	12135

OY	392	Leu1aVal111eProPromeLeuTyrAspAlaSerGlnArg1LeuYsPhe11aSn	411
Db	12135	-----	1213
OY	412	MetAenGlyLeuMetAlaAspPromeLeuValTyrYsAspArgGlnValMetAenMet	431
Db	12134	-----	1211
OY	432	TyrSerGlnGlnGlyYsGluThrPheArgGlyYsPheMetGln11eProLYsAenPhe	451
Db	12110	CGTACGCGCAGCCCGCAGAGACCGCGACCGCGTGCCTAGAGACATCC-----	1206
OY	452	GlyLeu11eAlaSerPheLeuGluArgYsThrVal11aGluCybVal11eLeuTyrTyr	471
Db	12059	-----	1201
OY	472	LeuThrYsLYsAenGluAenTyrLYsSerLeuVal11ArgArgSerTyrArgArg	489
Db	12017	-----	1196
OY	490	-----	501
Db	11963	CACACCGAACCGCGCGGACCGGTGAGGCCCGGTCCGAGAGACATCCGCGAGACAG	1190
OY	502	-----	510
Db	11903	CGCGTCCGCGTCCCGCATACGCGCGCAGCGCCGCGCAGACCGCTT-CACCTACTCTTAC	1184
OY	511	ProMetProArgSerSerGlnGlnGluYsAenPulYsGluYsGluYsGluValaGlu	530
Db	11844	CGAGGCGCCGAGACCGTTCCGAGAGCTCGGCGGTTGACGCGTTCCCGCGGCGGCC	1178
OY	531	LYs-----GluGlnGluYsProGluVal-----	538
Db	11784	AGCGCCGCGCCAGTCGCGCGTCCGAGAGCGTGGGCTCCGGTCCGCGACAGCGTCCCT	1172
OY	539	-----	552
Db	11724	CGCGGTCCGCGTCCAGACACCGCGCGAGTCCGCGCTTTTGGCGCGAGTGAACGCGCG	1166
OY	553	ThrSerGlyGluAenPheAenArgGluYsGluVala11aSerYsGlyArgYsThrAla	572
Db	11664	CTTCCGAGGCGGAGTGAATGAGCACTCCGCGCTCCCGACTGTCGCGCGGACACCC	1160
OY	573	AsnSer-----GlnGlyArgArgYsGlyArg11eThr-----	583
Db	11604	TCATCTGCACCGGACACACCGCGCGGCGGAGAACCAACGCGGCTCAACGCTACGCT	1154
OY	584	-----	595
Db	11544	CTTCCAGCGCCGCGAGCGCGTGGCGGCT-CCGCGTGCAGCGCGAACTTCAGAGAGGCC	1148
OY	596	-----	603
Db	11485	GTAACCGGAGCAACGTACCCCGGACACCGGACGTGCTCGTCAAGCATGCTGCTCGAG	1142
OY	604	Leu11aSerMetGluLeuAen-----GluSerSerArg-----	614
Db	11425	CCGCGACCGCGCGCGGAGACACCGCGTCAACCGAGTCGCGCAGGGGTAAACCGGCAACG	1136
OY	615	-----	618
Db	11305	GCGGGGAGATCAACGCGCCAGTAAGTGTGTGTAAAGGGGTAGGTGGGAGGAGTGGTG	1124
OY	618	-----	618
Db	11245	AGGTGACCTGGGGGAGAGCGCGGGTCCAGGTAGAGGTGGCGCGTCCAGGCGCTCG	1118
OY	619	GluMetCgluThrAlaYsLYsGlyLeuLeuGln11eGlyArgAenThrSerAla11eAla	638

```

Db      11185 GCGGTGAGAGTGCAGGCGGCTGTCTCTGCGGCGAGGCTCCGGTCCGGTCCGCA 11126
QY      639 ArgMetValGlySerIleThrValSerGlnCysIleAsnPheTyrPheAsnTyrIleVal 658
Db      11125 TGGACACTCCGGTCCGGGCGCTCCGGCTGTCTC-AATGGCATGGACAGTACCGGGTG 11067
QY      659 ArgGlnAsnLeuAspGln-----IleLeuGlnGln-HisIleLeuIleValMetGln 674
Db      11066 GGCACCTGACCTCGATGACCGGACGGTGGCGCTCTCCAAACAAACCCAGCACCGCGCGA 11007
QY      674 uIySGlUaGAsnAlaArgArgIleValIySylValaProAlaIaIaIaSerGlnUaI 694
Db      11006 GAACCTCAACCCGGTTCGCGAGATTACGAAACCAATACCCCGCATCGAAAAACCGCTCGCC 10947
QY      694 aAlaPheProProValIaGluAspGluUaIaSerGluValaSerGluValaSerGluVala 714
Db      10946 CGGCTGTCTCTCCGGGACCGGTG---GAACACACCGGACCGCGGACCTCGGGGGGGGAT 10890
QY      714 uGluGluMetValGluGluUaIaGluAlaLeuHisAlaSer----- 727
Db      10889 ACCGGACACACCGCGCTCCAACTCGCCCTTCAGGTCTCTCAAGTCCCATGAGGAGTGAAGGC 10830
QY      728 ----GlyAsnGluValaProArgGlyIleCysSerGly----- 738
Db      10829 GTAGTTCACCGGTATCAACCGAGCC---TGCACACCTCCGCTCGACAGCGGCCACATA 10773
QY      739 -ProAlaThrValaAsnAsnSerSerSerAspThr----- 748
Db      10772 ACCGGCACCGCGCGCGCATCCCGGAAACACCGGTGACCGCGGACCGGTACCGCGGC 10713
QY      749 -----GluSer-IleProSerProHisThrGluAlaIaIaVala- 760
Db      10712 TACTGTACTCCGGGCGCACTTCGCGCGCGCTCGACCTCGCGCGGACGAGCGGAAC 10653
QY      761 -----AspThrGlyGluAsn----- 765
Db      10652 AGCGCGCATCGACCAACGACCGCGCATATCAACCGCGCGGACGAGCGGAC 10593
QY      766 ----GlyProIySPProAlaThrLeu----- 773
Db      10592 CACAGCGCGCGCATCTCCAGGCAACACCGCCCGGACCGGTGGACGAGCATCTCCCG 10533
QY      774 --GlyAlaAspGlyProPro----- 783
Db      10532 CTGGGAATAGCCACACCGGACGACGTCACACCATCGCTGCGAGTACGGGCGCAG 10473
QY      783 roThrProProArgArgThrSerArgAla----- 792
Db      10472 CGACACCATCAACCGCGCGTCACGGGCTGGACACCATCCACCGCTCAACCGCGGCCCC 10413
QY      793 -----ProIleGluProThrPro----- 798
Db      10412 GCCCGACACCACTCCAGCAAAATCCCACTCAATGAACACTCAGGCGCGGCGACACTC 10353
QY      798 ----- 798
Db      10352 CCGCATCGACGCGGCAAAACCTCCGAAGACCCGACAGCGCGGACCATCCCAACCCA 10293
QY      799 --AlaSerGluAlaThrGlyAlaProThrPro-----ProProAlaPro- 812
Db      10292 CTGGCTCCCTTGAACAGGAAAAACAACCAACGACCGCCCAACCGGACCCGCTCTCCG 10233
QY      813 -----ProSerProSerAlaProProProVal----- 821
Db      10232 GTTCACACCGGACCGGTTACACACCCGCGTCCGATCAACCCCGGCAACCGCGCAA 10173
QY      822 -----ValProIySGluGluIySGluGluGluThrAla 833
Db      10172 AGCGCGGCAAGTCAAGCTCAACCGCGAGCGACCAACCGCGCGGATACCAAAACCGCGA 10113
QY      833 laAlaProProValGluGluGluGluGluGluIySGluGluIySGluGluIySGluGluIy 853

```

```

Db      10112 CCGGCGCGGCA-----TGTTCACCGGCACATCCCGCAGCGCCCAACCGG-GCC- 10064
QY      853 aAlaSPThrGlyIyValaGluGluProValIySerGluCysThr-----GluGluAlaG 871
Db      10063 --TTCACCACTCCCGCAACCGGCGCGAGCTCCCGGACACCGCGCGCAACCGCGCG 10006
QY      871 uGluGluIyProAlaIySGluIySGluAlaGluAlaGluAlaThrAlaGluIyVala 891
Db      10005 AAACCAACACCGGACCAACACCGACGAGCTGTGCTCTTCGGAGACTGCCAGAGAGCTT 9946
QY      891 eu-----LysAlaGluIySGluIySGluIySGluIyValaArgAlaThrThrAlaIySers 909
Db      9945 TCGCACCTCCCGCTCCGCTGGGTGGGTGGGGGCTTCAGAGGCTTCGGTGAAGCTGTGAGCT 9886
QY      909 eRGlyAlaProGluIySGluIySGluIySGluIySGluIySGluIySGluIySGluIySGlu 929
Db      9885 CGGGGGCTTCGGGGGCTTCGGGGGCTTCGGGGGCTTCGGGGGCTTCGGGGGCTTCGGGGG 9826
QY      929 laGluGlyGlyAspIySGluIySGluIySGluIySGluIySGluIySGluIySGluIySGlu 948
Db      9825 CTTCGGGGGCTTCGAGGATCAAGATGAGCGGTGTACCA-----CTGACACCGA 9778
QY      948 hrGlyAspProArgAlaAsnAlaSerProGluIySGluIySGluIySGluIySGluIySGlu 968
Db      9777 A-GCGGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 9719
QY      968 laArg-----AlaAlaAlaIle----- 973
Db      9718 AGCAACCGCACACACGACGACCGACGATCAATCCGCTCGGCTCGCATCGACATGCAAC 9659
QY      974 -----ProIleGluIyValaThrIyVala-----HisGluProPro- 985
Db      9658 GTCCGGCGCAACCCACCATACC-GCATGCGCAACCATCTTGATCAACACGACCAACCC 9600
QY      985 ----- 985
Db      9599 CCGACCGCGTGAAGATGACCAATATGACTTCAACGACCCCAACACACCGGACACTC 9540
QY      986 --ArgGluAspAlaIaProThrIySGluIySGluIySGluIySGluIySGluIySGluIy 999
Db      9539 AGCATCAAGATCCCGCCCATAGTCCCAACCGCTTCGCTCGATCGATGATCAACCGAG 9480
QY      1000 -----ProProProGluIySGluIySGluIySGluIySGluIySGluIySGluIy 1013
Db      9479 CCGGCTCCCGTCCATGCGCTTCACACCATCAATCCGCGACCGTACCGG---- 9424
QY      1013 lngIySGluIySGluIySGluIySGluIySGluIySGluIySGluIySGluIySGluIy 1033
Db      9423 -----CATTCGCGAAG-----CTGAGATTCACCGCTCTCGGACGACCACT 9378
QY      1033 luAlaPheAlaIaGluAlaGluIySGluIySGluIySGluIySGluIySGluIySGluIy 1053
Db      9377 CGGCGCGCGTCAAAACATTCGACGACCGCTCGATTCACCGCGGACCAACCA----- 9325
QY      1053 euProPheProValaProProAlaGluValaIleValaIaSerProHisAlaProAspPro 1073
Db      9324 --CAACCGGCAACCGGAT-----GCCCAACCGCGGCGATCCGA 9285
QY      1073 eAlaIaPheSerTyrAlaProProGluIySGluIySGluIySGluIySGluIySGluIySGlu 1093
Db      9284 CAACCGCTCCACGACACACCGCGGACCTTCGCGCGGACGAGGCTCCAT----- 9235
QY      1093 rgProValaIySGluIySGluIySGluIySGluIySGluIySGluIySGluIySGluIySGlu 1113
Db      9234 --CGGCGCGATCCGGAACGCTTCACCGCGCATTCGCGCGCAACCGCGCGCGCA 9177
QY      1113 yHisProSerValaIySGluIySGluIySGluIySGluIySGluIySGluIySGluIySGlu 1133
Db      9176 AAATCCACGA-----ACAATCCGGGTGGCGGCGCATCCGTAC 9138
QY      1133 euHisValaProTyrSerGluIySGluIySGluIySGluIySGluIySGluIySGluIySGlu 1153
Db      9137 ACCACCGGCAACCGCATGTCACTCGCACCGCGCAACGCTCGG----- 9091

```

QY 1153 roLeuProMetAspProLyLeuLeuAlaProPheSerGlyValIyGlnGlnLeuS 1173
Db 9090 -----CCGGTTGTGTGACAGGCGCACACGACGACGACGACGCGGTGC 9048
QY 1173 erProAlaGlyGlnAlaGlyPro-ProGlnSerLeuGlyValProThrAlaGlnGlnAla 1192
Db 9047 GATGGAGAGGAGGAGCCCTCCAGCCCGGTACAGATGCGCGGAGAGAGCGCT 8988
QY 1193 SerValLeuArgGlyThrAlaLeuGlySerVal-----ProGly 1205
Db 8987 CGCGACCTTGGCGGTGACAGATAGCTTGTACTCTCCGGGGGCTCGTGCAGCCGGA 8928
QY 1206 GlySerIleThrLySerGlyIleProSerThrArgValProSerAspSerAlaIleThrTyr 1225
Db 8927 GGGGTACTCTCGATGGGTCAAGCCGAC-----GTATAC 8895
QY 1226 ArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrGlyThrIleThrArg 1245
Db 8894 CCGCGTCCG-----GCTGCCGTGCACGGA 8871
QY 1246 IleIleGlyGlnAspSerProSerArgLeuAspArgGlyValArgIleAspSerLeuProLyS 1265
Db 8870 GTC-----CGGGTGCATACCGCGCGCTCGATCCGCTCCACGA 8832
QY 1266 Gly-----HisVal-IleTyrGlnGlyLyValSerGlyHisValIle 1278
Db 8831 GGTCTCCAGAACACGCGCGTCTCGGATCTCATTCGCCACCGCGCGCGAAATCCC 8772
QY 1278 userTyrGlnGlyGlyMetSerValThrGlnCySerGlyGlnAspIlyArgSer-SerS 1298
Db 8771 GAAGAACTCCGCGTCAACTCCGCGCGGTGAGAAACCCCGCTCCGACATATGT 8712
QY 1298 erGlyProProHisGlyThrAlaAlaProLyAspGlyTyrArgMetMetGlnGlyArgV 1318
Db 8711 CTTCGCCGACACCGCGCTCGGCTGTAGACGGAATCCAGGTCCACGCGCGGTGCT 8652
QY 1318 alGlyArgAlaIle-----SerSerAlaSerIle----- 1327
Db 8651 GGGCGGCGCGCTATGACGTACGCTCTCGGTACACAGCTTCCACAGCGCGCTCAGGGT 8592
QY 1328 -----GlnGlyLeuMetGlyArgAlaIleProProGlnArgHisSerP 1342
Db 8591 GCCCGCCACCGGGGAAACGGACGCGCATGCGGACGATGACGCGGCTGTCTGGC 8532
QY 1342 roHisHisLeuLyGlnGlnHisIleArgGlySerIleThrGlnGlyIleProArgS 1362
Db 8531 CACGG----- 8527
QY 1362 erTyrValGlnAlaGlnIleAspTyrLeuArgGlnAlaLyLeuLeuLyAspGlnG 1382
Db 8526 -----CGACGGGACGTGGCGGCTCATCGCGCGG-G 8497
QY 1382 LyThrProProProProProSerArgArgLeuThrGlnAlaIlyLyThrGlnAlaI 1402
Db 8496 GCCCGGCGCGCTCGCGCGCTGACGCGGTGCGACGACGTGCGGCGCGGTGCGG 8437
QY 1402 euGlyProLeuLyLeuLyProAlaHisGlnIleValIleThrValLyGlnAlaG 1422
Db 8436 TGGGGTGTGCAAGATG-----ACGGTGGCGGTAGCT 8404
QY 1422 LyArgSerIleHisGlnIleProArg-----GlnGlyLeuArgHisIleThrProG 1438
Db 8403 TCGAGCGCTGTGCGGATGAGCGGTGCGGAGTTCCAGGAGGTGAGGATCGAAAC 8344
QY 1438 IuLeuProLeuAlaProArgProLeuLyGlnGlySerIle-----ThrG 1453
Db 8343 CGCGCACTTGAACGCGCGCTTCCGCGCGATGCGCTGTGTTCCGATGCCGACGACAA 8284
QY 1453 InGlyThrProLeuLyTyrAspThrGlyAlaSerThrThrGlySerLyLeuHisAspV 1473
Db 8283 AGGCGCGCTCGGTACGG-----ACGAGTTGACCACTTCCGAGCGGTTCGG 8236

QY 1473 al-----ArgSerLeuIleGlySerProGlyArgThrPhePro 1486
Db 8235 TCTCGTTACAGATGACAAACCGCGCGCGAGCTCGGCGCGCGCTCGCGGTGTC----- 8183
QY 1486 roValHisProLeuAspValMetAlaAspAlaArgAlaLeuGlnAlaCysTyrGlnG 1506
Db 8182 -----GGGAGCTGTGTCCGCTCGTCAGAGGCGCTCCGCGTCCGCGCACGG 8134
QY 1506 IuSerLeuLySerArgProGly----- 1513
Db 8133 TGTCTG---AACAGCGCGCGGAGCGATGGCGCTTCTCGCGGCGGAAACGCTCCCACT 8077
QY 1514 -----ThrAlaS 1516
Db 8076 CGAAGTCAGACAGATGACGACACGCTGTTCTGCGCCATGGGCGCCAGAGCGCGTCA 8017
QY 1516 erSerSerGlyLySerIleAlaArgGly-----AlaP 1527
Db 8016 CGGCGCGTGTGCGGTCTCATGCGCGCGAGCGCTCGCGCGACAGACTCTCTCGCGCGCAC 7957
QY 1527 roValIleValProGlnLeuGlyLyLeuProArgGlnSerProLeuThrTyrGlnAspHisG 1547
Db 7956 CCGCGCGCATGCTT-----CCCGCGCCAAAGGCGCCAGGCGACGAGGTGCGG 7906
QY 1547 LyAlaProPheAla----- 1551
Db 7905 GCAGTCCGCGCGGACGCGCGACGCTCGCGCGAGCGGTGACGCGCGGTGCGCGCGCT 7846
QY 1552 -----GlyHisLeuProArgGlySerProValThrMetArgGluP 1565
Db 7845 ACGCACCTGCGCGCGGTGCGCGCGATGTGCGGTGACGAGGAGAGAGACGAAAGCGCT 7786
QY 1565 roThrProArgLeuGlnGlySerLeuSerSerIleValAspGlnAspArgIlyS 1585
Db 7785 CCAACCTTTTATGTCCGAGTCAGTGTGTGACAGACTTCGCGCGCGAC----- 7736
QY 1585 euThrSerThrProArgGlnIleAlaLyS-SerPro----- 1596
Db 7735 --ACCTTGGCGCGCGGACGCGTCAAGGTCTCGGTGACAGCGGTGCGATCAGCGCT 7678
QY 1596 ----- 1596
Db 7677 CGTCCAGAAATCCGCGCGGTGTGAGAGACGCGCTTCCGCGGATACGCGGTGACAGCGCG 7618
QY 1597 -----HiserThrValProGlnHisHisProHisProIleSerProTyrGlnHisLeu 1614
Db 7617 CGAGCGCATCAGCTCGCGCGCATGCTCAGAGCGCGCT--GCACGACCTGCGACGCAATGTC 7561
QY 1615 Leu----- 1615
Db 7560 CTTCGACGTTCTCGCGGAGTTCGCGCGCGCGCGCGCTCGGTCCGCGACGCGTGTGA 7501
QY 1616 -----ArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAla 1631
Db 7500 GACACGAGGCTTCGCT--CGCTCGCGCGCGCGAGCTTCGCGCGCGCGCGCGCGAGTGA 7442
QY 1632 PheAspPro-----ThrSerIleProArgGlyIleProLeuAspAlaAlaAla 1648
Db 7441 CCCATGCGCGCGGTATGACACCGTGCAGACGCGCTGCCAG--GAGGCGGCGCGCGCT 7385
QY 1649 TyrTyrLeuProArgHisLeuAlaProAspProThrTyrProHisValLeuTyrProProTyr 1668
Db 7384 GCGCGCGCG--CGCGGACAC-----CGCGCGCGCGGTAGATCCC----- 7350
QY 1669 LeuIleArgGlyTyrProAspThrAlaAlaLeuGlnAspArgIleThrIleIleAspAsp 1688
Db 7349 -----CGAGGC----- 7344
QY 1689 TyrIleThrSerGlnIleMetHisIleAspThrAlaThrAlaMetAlaGlnArgAlaAsp 1708
Db 7343 -----CGGACGCGGACGCTGTCTCGCGTGTGCGCGCGCGCGGAGAC 7302
QY 1709 -----MetLeuArgGlyLeuSerProArgGlnSerSerIleAlaLeuAspntYrAla 1725

Db 7301 GCGCGAGCAGAGCTCAAGGCGCGTCGCT-TCGAGCTCTCGGACAGCTCTCCGGGGCTCT 7243
QY 1726 Alagly-----ProArglyLysleuSer 1735
Db 7242 GTGGGGCTGTGAGTCTCCGGCTCCGGTCAAGCTCTCTCCGGAGTTCATCA 7183
QY 1736 GlnValProHisLeuPro--ValLeuValProProThrProGlyThrProAlaThrAla 1754
Db 7182 AGCGCGCCAGATATCCGGGAGTTCAAGGGCGGCACTCGTCCGAGCCCAACATCGG 7123
QY 1755 MetAspArgLeuAlaArgLeuProThrAlaProGlnProPheSerArgHisSerSer 1774
Db 7122 CCCCCCGCTCTGGCGCTCTCCCGGAAACAAGGCCACCGCCCTCCGGTCAAGCGCC 7063
QY 1775 SerProLeuSerProGly----- 1780
Db 7062 AGACCTTGCGCGGCGGAGACCGCGATGTCGAGACCGCTGAGAGAGCCATTCAGCTCC 7003
QY 1781 -----GlyProThr 1783
Db 7002 TGACCGCGTCCGAGAGGACACAGACGAGAGACGAGTGTCTCTCC 6943
QY 1784 HisLeuThr----- 1786
Db 6942 CCGCTGACTCATGAGAGGCTTGAGCATTCGCGCATTCGCGACCGCGCGCCGATG 6883
QY 1787 -----LysProThrThrSerSerArgLysArgLys 1797
Db 6882 GCACCTCAAGGCGGCGGAGAGGACCGCGCTCGAGCAACCGTTGTAGTCCGTCGA 6823
QY 1798 ArgAspArg-----AspArgLysArgAspArg 1806
Db 6822 GAGGCTCGCGCTCGGTCGCGGCGGTCAACAGAGAGGACCGATGAGTGGGCG 6763
QY 1807 AspArgLysArgLysSerLysLeuThrSer--ThrThrValGlnHisAlaPro 1825
Db 6762 GAGCGGGAGGATTCGTCAGGCTTTCAGGTGACGCGTAGCGCCAGGCAATCGG-- 6706
QY 1826 IleTrp----- 1827
Db 6705 --TGGGGCTTGGGTGTCTGTCGCGGGGTGGGGGTGTGTGAGCCAGTAATGGT 6649
QY 1827 ----- 1827
Db 6648 GGTGTTGAAGGGGTAGTGGGAGGTGGTGTAGTGGCTTGAGGAGCGGGGTCC 6589
QY 1828 -----ArgProGlyThrGlnLysSerGlySerGly 1839
Db 6588 AGGTGAGGTGGCGCGTGGGCGCCAGGCTCGCGGTGGAGTCAAGCGCGTGGCGG 6529
QY 1840 SerSerGlyGlyGlyGlySer-----SerSerArgPro 1851
Db 6528 TGTGCTCTTGCGCGGAGGTCCCGGTGATGACATCTCGGTCGCGGCTCGGCGG 6469
QY 1852 AlaSerHisSerHisAla----- 1857
Db 6468 TGTGCTCATTCGATGAGAGTACCGGGTGGCACTGACCTGATGAAACCGAGTGG 6409
QY 1858 -----HisGln-His-----Se 1861
Db 6408 CCTCTCAACAACAACCGACCGAGAGAACTCAACCCGTTCCGAGATTACGA 6349
QY 1861 rProIle----- 1863
Db 6348 ACCAATAACCCCGCATGAACCGGCTCGCGCGTGTCTCCGCGACCGTGAACA 6289
QY 1864 ----SerProArgThr-----GlnAspAlaLeuGlnGlnArgProse 1876
Db 6288 CCGGACACCGCGGACTCGCGGGCGGATACCGGACAGACCGGCTCAATCGCCCTCA 6229
QY 1876 rValLeuHisAsnThrGlyMetLysGlyLysLeuThrAlaValGlnProSer----- 1893

Db 6228 GGTCTCCATGCGGGAGTGAAGGCGTAETCCACCGGATTCACAGAGCTGCACAC 6169
QY 1894 -----LysProThrValLeuArgSerThr-- 1901
Db 6168 CCTCGGCTGACAGACGGGACATTAACGGGCCACCGCGCGCGATGCC--CGAACAACAC 6110
QY 1902 -----SerThrSerSerProValArgTr 1909
Db 6109 GTGAGCGCGGACCGTTGACCGCGGACCTTCACACCGGCGGACTTGGCGCGCGG 6050
QY 1909 o-----AlaAlaThrPheProProAlaThrHisCysProLeuGlyLys 1924
Db 6049 TGACCTCGCGCGGACGAGCGGAAACAGCGGCTATGACACAGACCG--CGATATC 5993
QY 1924 rLeuAspGlyValTyPro-----ThrLeuWe 1933
Db 5992 CGGCAATACACCGCGCGGAGAGGACACAGCGCGCGCATCTCCAGGACACAGCG 5933
QY 1933 rGlnProValLeuLeuProLysGlnAlaProArgValAlaArgProGlyArg-- 1952
Db 5932 CCGGACCGGTGGCAGACAGCGATCTCCCTGGAAATGACCAACACCGGACGACGTC 5873
QY 1953 -----AlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerG 1968
Db 5872 AACCCATGCGCTGCGATGACGGGCGACGACACCATACCGCCACGTCAGGGCT-- 5815
QY 5814 -----GACACCATTCACCGGCTCCACCGGCGCGCGCGCGACACACCTTCAC 5765
QY 1988 Lser-----GlyHisAlaThrLysAlaArgThrProAlaLysAsn-- 2001
Db 5764 AATCCCACTCCACATGACACTGACGCGCGCGGACACTCCGATCGACCGCGAAC 5705
QY 2001 ----- 2001
Db 5704 ACCTCGAAGACCCAGACGCGCGGACCCCATCCCATCGGTCGACCGACAGGG 5645
QY 2001 ----- 2001
Db 5644 AAAACAACAACGCGCCCAACCGGACCCCGTCTCGGGTCCACACGACCGGTG 5585
QY 2001 ----- 2001
Db 5584 ACCACACCGCGTCCGATACCCCCCGCAACGCCCGCAACGCCCGCAAGTCAAGC 5525
QY 2001 ----- 2001
Db 5524 TCACCCCGACGACCAACCGCCGATACCAACCGGACCGCGCGCATCTGCACC 5465
QY 2002 -----LeuAlaProHisHisAlaSerProAspProPro-- 2012
Db 5464 GCCATATCGGACCGCCACACACCGGCTCCACACACTTCGCAACCGGCGCGGAGC 5405
QY 2013 -----AlaProProAlaSerAlaSerAspPro-----His 2022
Db 5404 TCCCGCACCAACCGCGGAGCGCCCGCAACACACACCGGACACACACTGCGGGA 5345
QY 2022 sArg-----GlnLysThr-----GlnSerLysProPheSerLysGlnGln 2037
Db 5344 TGAGGACCGGCACTCCGACACTTCAGAGCGGAGCGGAGTCTGTGGCGCTCTCGGC 5285
QY 2037 uLeuArgSerLeuGlyTyHisGlySerSer----- 2047
Db 5284 TCGCGGTGTGGGGGCTTCAGAGATCAATGAGGCGTTGGTACACTGACCAACAGGG 5225
QY 2048 -----TySerProGlyLysValGlnProValSerProValSerProSerLe 2064
Db 5224 GAGACCGCGCGGACGCGCGGCTCAACCGCGGCACTCCCGCGCTCGGTCAACAAC 5165
QY 2064 uThrHisAspLysGlyLeuProLysHisLeu--GlnGlnLeuAspLysSerHisLeuGln 2084
Db 5164 CACACACCAACGAGCGGACGATGCCATGCGCGGCTCGGCTC-----ATCCATATCA 5112

```

Qy 2084 lylguleuAArgProLysGlnProGlyProValLysLeuGlyGlyAlaAlaHisLeuP 2104
Db 5111 -----CGTCCGGGGAACCA-----CCCATACCGCATCGCCAAACACATCTTG 5069
Qy 2104 roHlSleuAArgProLeuProGlnSerGlnProSerSerSerProLeuGlnThrAlaP 2124
Db 5068 ATCCACAC-CAGCCCAACCCCGCGCTGAGCATGACCAATATTGACTTCAACGACCC 5010
Qy 2124 roGlyValLysGlyHisGlnArgValThrLeuAlaGlnHisLisSerLiuVallet 2144
Db 5009 CA-----ACCA 5004
Qy 2144 hrglnAerTyThrArgHisHisArgProGlnGlnLeuSerAlaProLeuProAlaProLeuT 2164
Db 5003 CACCGGACGGCCACCATCAGATCCGCCCACTAGCTCCCGACGACCGCTCGCTGAT 4944
Qy 2164 ySerPheProGlyAlaSer-----CyProValLysAerLeuArgArgProProS 2181
Db 4943 CGGATCACCCAGCGCTCCCGCTCCATGCGCTCCACCATCCATCCGCGCACCT 4884
Qy 2181 eHApLeuTyThrLeuProProPro-----AerHisGlyAlaP 2193
Db 4883 CAACCGGCAATTCCTCCCAACGCTTACCAATCACCGCTCTGCGACGACCATCGGCGC 4824
Qy 2193 roAlaArgLysSerProHisSerGlnGlyGly-----LysArgSerProGlnProAln 2211
Db 4823 CGTCAACCATTCGACGACCGCTCTGATTCACCGCGGACCCACACACACCGCCACAC 4764
Qy 2211 yHrSerValLysGlyGlyGlyLysAerGlyLysGlnProValSerProProGlu 2229
Db 4763 CGGATGCCCAACCGCGCGCATCCGACCAACCGCTCCACGACGACCGCGCACCTC 4704
Qy 2230 -----GlyMetThrGlnProGlyHisSerArgSerAlaValTyProLeuLeuTyArg 2248
Db 4703 GGGCCAAAGCGGTCCATCCCGCTCCCGCTCCGCAAC-GCC-----TTGACCGGC 4657
Qy 2248 spGlyGlnGlnThrGlnProSerArgMetGlySerLysSer-----ProGlyAer 2265
Db 4656 CATCGCGCGGCAACCCCGCTCGCGCAAACTCCAGAACTCCCGCGCGCATCA 4597
Qy 2265 hrSerGlnProProAlaPhePheSerLysLeuThrGlnSerAerSerAlaMetValLys 2285
Db 4596 CGGTCAACACCGCGCC-----AACGGCAT-GTCACACTCG 4562
Qy 2285 erTybLysGlnGlnLysAerLysLeuAerThrHisAerAlaG 2299
Db 4561 CCAACCCCGCAACGCTGACCGCGCATGCAACGCGCAACGACGACGACCGGTA 4502
Qy 2300 -----AerGlnProGlyTyAer 2306
Db 4501 TCACGCGTACGCGCGGACCTCCAGCCCAAGGTGAGGACACACCGCGACACCA 4442
Qy 2306 leSerGlnProGlyThrGlnLysAerAerProAlaLysThrGlyThrGlyLeuMet 2326
Db 4441 CTCGCCGACGATACCGGTGAGGAGTAGCGTCCGCTTCGCTCGGCTCGGCG-----G 4388
Qy 2326 hrTyArgSerGlnAlaValGlnGlnHisLisSerThrAerMetGlyLeuGlnAlaLys 2346
Db 4387 AGCGGAGGTCTGATTCCTGCGCATCACGCGCGGTGAGCGCGGCTCGCTGCG 4328
Qy 2346 leArgValAlaLeuMetGlyLysTyArgGlnThrGlnGlnSerProProLeuSerAla 2366
Db 4327 AGCGAGT-GGG-----GTCGATACCGCGCC-----GTCGATGCGCTCCGAGAGG----- 4282
Qy 2366 snAlaPheAerProLeuAerAlaSerAlaSerLeuProAlaAlaMetProLysThrAla 2386
Db 4281 -----TCTCAGACAAACGCGGTGCTGCGATCATCGACCGCGCTCAGCGCGCA 4230
Qy 2386 laAerGlyAerSerAerPheThrLeuThrSerProGlyGlyGlyLysAlaValLys 2406
Db 4229 AATCCGAGAGACTCCGACATGA--ACTCCGCGG-----CATGATGACGAAACC 4182

```

```

Qy 2406 erGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAer 2426
Db 4181 CCCCTCCGCAATACACTTCCCGGACACCGCGCTCGGGTCTGACAAACC----- 4129
Qy 2426 rGProProSerValSerSerValHisSerGlyAerCybAerArgThrProLeuT 2446
Db 4128 --CCTCCACATCCGACGACGATCACCGGAAACCGGGAAACCGCTCCACCCCT---- 4075
Qy 2446 hrAerArgValThrGlnAerArgProSerSerAlaGlySerThrProPheProTyAer 2466
Db 4074 -----CCACACCAACCGCCACACGCT--CCGCGAAC 4041
Qy 2466 roLeuLeuAerArgLeuGlnAlaGlyValMetAlaSerProPro-----ProProGlyL 2484
Db 4040 CA-----CCCAACCGGGAACCGGACCCCAATACCAACGCGCATCGGCTCTGCGC 3987
Qy 2484 euProAlaGlySerGlyProLeuAlaGly 2493
Db 3986 TTCGAGCTCTCTGCGCGCGTCCGAGGA 3958

```

RESULT 15

```

US-08-804-227C-13/C
; Sequence 13, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kunze, Stuart A.
; APPLICANT: Roebuck, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYMERIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII (DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804.227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13987 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..13987
; US-08-804-227C-13

```

Alignment Scores:

```

Pred. No.: 1, 01e-20 Length: 13987
Score: 652.00 Matches: 648
Percent Similarity: 30.21% Conservative: 260
Best Local Similarity: 21.56% Mismatch: 1034
Query Match: 4.93% Indels: 1081
Gaps: 133

```

US-09-522-753-5 (1-2517) x US-08-804-227C-13 (1-13987)

QY 15 ThrGluProArgTyProProHis-----SerLeuSer 25
|||:::
10684 ACCCGAGACACGACACCCACCCCATCCGGAAACGACCGCTCCCGGCAATGAG 10625

QY 26 TyrProVal-----GlnIleAlaArgThiHisThrAspValGlyLeuGlu 41
|||:::
10624 CACCCAAATACCCCGATCCCGCACTCCCGAAACCAACCCGACACCCGCA 10565

QY 42 TyrGlnHisHisSerIleAspArgPyrAlaSerHisLeu-----SerProGlySer 57
|||:::
10564 CACCAACCGACACCGCGATCCCGCACTCCCGAAACCAACCCGACACCCGCA 10505

QY 58 IleIleGlnProGlnArgTy-----ArgProSerLeuSerGluPheGln 73
|||:::
10504 CTC-----CCCCAACCGCGCTCCATCAACACCGAAATGAAACCCATGGGACATCCA 10451

QY 74 Pro-----GlyAsnGluArgSerGlnGluLeuHis 83
|||:::
10450 CCGCCGGCACTCCACCCCATCCCATCAGCGTGAACAACCGACTCCGACACACCCAC 10391

QY 84 LeuArgProGluSerHisSerTyLeuProGluLeuGlyLysSerGluMetGluPheIle 103
|||:::
10390 ATCAACCGCAACACCAACCGACCGCGCGCC----- 10361

QY 104 GluSerIleArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProLeu 123
|||:::
10360 ATTCAACCGCGACACCAACCGCACTCC--CCAACCCCTCAACAACCCCGGACACCG 10302

QY 124 LeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGly 143
|||:::
10301 ACTCCGACCGCCCGACCAACCATCCCGCAACCGGCAACCCACCATCAACCC--- 10245

QY 144 LysLeuGluProValSerProProSer-----ProPro 154
|||:::
10244 -----CACCCGCGCCACCAACCGCAACCGCATACCCCAACAACACCCCGGCA 10191

QY 155 HisThrAsp-----ProGluLeuGluLeuValProPro-ArgLeuSerLysGln 170
|||:::
10190 CATACGACGACGACCTCCCGCAACCGAATGACCAACACCGACGACCTCCACACCC 10131

QY 170 uGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGluGlnGlnIle 190
|||:::
10130 GAGCTC---CAACGCGCGGAACAA-----CGCACTCTCAACGCAACCAACGAGGCT 10080

QY 190 eSerIleLysLeuLysLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 207
|||:::
10079 GAGCAAACTCCGTCGACCAACCAACCCCAAAACCGAATGACAT--CCGGAACATC 10021

QY 208 -----ProGluProGluLysProValSerPro----- 216
|||:::
10020 ACATCCCGCAACGACGACCATCTCCCGCAACCGCAACCGCAACCTCATCCAGCAC 9961

QY 217 -----ProPro-IleGlu---SerLysHisArgSerLeuValGlnIleIle 230
|||:::
9960 TCCGCAAAACACCCCAACCCCGCATCAACCAACCGCCCATACCAACCACTCGGACCC 9901

QY 230 IeTyAspGluAsnArgLysLysAlaGluAlaIleHisArgIleLeuGluGluGlyP 250
|||:::
9900 TGAACCGGTAAACAAACCCCGCAACCGCAACCGCGCGC----- 9860

QY 250 roGlnValGluLeuProLeuTyArgGlnProSerAspThrArgGlnTyHis----- 267
|||:::
9859 -----ACTCCCGCGACACCAACGACCGCGCCCGGAAGACACACCCAGAGACC 9811

QY 268 -----GluAsnIleLysIleAsnGln-----AlaMetArgLysL 279
|||:::
9810 CCGGACGACACCATCAAGACCGGCAAGCACTCTCAAGCGCCCGGCAACCAACGACC 9751

QY 279 yLeuLeuIleuTyRheLysArgArgAsnHisAlaArgLysGlnTyrLysGlnLysPheC 299
|||:::
9750 GCAGCTCAACCGAAGCAACGACGCAACCAACCAACGA----- 9713

QY 299 ySGlnArgTyAspGlnLeuMetGluAlaLeuGluLysValGluArgIleGluAsnA 319
|||:::
9712 -----CCCCGCAATCCCGGATCCGACCC 9685

QY 319 snProArgArgArgAlaLysGluSerLysValArgGluTyrTyroGluLysGlnPheProG 339
|||:::
9684 ACCCGCTCCACCACTCAACCGAACCGCAC-----C 9655

QY 339 LuIleArgLysGlnArgGluLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 359
|||:::
9654 GCTCGCGCCGCAACGCGCCCTCCGACCGGAGAAAGAC----- 9614

QY 359 eArgLysLeuSerMetSerAlaAlaArgSerGlnHisGluValSerGluIleIleAspGlyL 379
|||:::
9613 -----CCAGCAACCGACCGCTCAACCGGCTCGGCGGCTC----- 9578

QY 379 euSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProMetL 399
|||:::
9577 -----AGATTCAAGGAGCGTCCCGCTGCTCGCTCCGCGCC 9544

QY 399 euTyAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspP 419
|||:::
9543 TC----- 9542

QY 419 IometLysValTyLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluT 439
|||:::
9542 ----- 9542

QY 439 hrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAspPheLeuG 459
|||:::
9541 -----CGCGGCGCTTCTCCAG-----GACCAATGGGCGGTGGTCCCGCTCAACCCG 9493

QY 459 LuArgLys-ThrValAlaGluCysValLeuTyTyTyLeuThrLysLysAsnGluAsn 478
|||:::
9492 AACGCGCAACCGCGCGCCG-----CGACCGGCTCCACCGCGCGC 9451

QY 479 TyLysSerLeuValArgArgSerTyArgArgArgGlyLysSerGlnGlnGlnGln 498
|||:::
9450 GGCACGACCGGCGTCCGGTCAAGAG--CTCAACCCGACCGGAAGCCCACTCCACCTTCCA 9392

QY 499 Gln 518
|||:::
9391 CACCGCGCATCAATGATGCGCGCAACGACCATCGCATCCGATCCCTGCAC-- 9335

QY 519 GluLysAspGluLysGluLysGluLysGluLysGluLysGluLysGluLysPheGluVal 538
|||:::
9334 -----CACCTTGATCAACCAACCAACCGCGCGCTGCGCATGCGCATGTT 9284

QY 539 GluAsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyLysAsnAsn 558
|||:::
9283 CCACTTCAACGA----- 9272

QY 559 AspGluLysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnLysArgArg 578
|||:::
9271 CCGGACGCAACGAGATCGCCACCGCTCCCGCGCGTCACTGGCAAGGACGACCCG 9212

QY 579 LysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluLysAlaIleThrPro 598
|||:::
9211 CTCGATCGATCAACCAAGCGCGCTCCCGTACCGTCCGCTCCACAC----- 9164

QY 599 GlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGlu 618
|||:::
9163 -----ATCCAGCTCGCGGAGCAACGCGCGCATCGAGCGCTCAACGATCAAC 9113

QY 619 GluMetGluThrAlaLysLysGlyLeuLeuGlnHisGlyArgAsnTrpSerAlaIleAla 638
|||:::
9112 CCGCTG-----CTGCGCGGACCACTGGCGCGCTCAACCGCATTTGCT 9071

QY 639 ArgMetValGlySerLysThrValSerGlnCysLysAsnPheTyRheAsnTyRylsLys 658
|||:::
9070 CCGACCGTC----- 9062

[illegible]

QY	1515	AserserserClgylserllealaaagly-	-----	1525
Db	6234	CAGTAGC-----GATCGCGTTGMAAGGGTAGGTGGAGAGTCATGCCCGCGGTG	6181	
QY	1526	-----	AlaProvalilleValProgluleuGl	1534
Db	6180	CTCCCGGTGTGAGATGCCGTGGCGTCCAGCGCGCGTCCGGGTGMAACACCTGGCCAGC	6121	
QY	1534	YLYsProArgrglInserProleuthrTYrglUasrphIsolYalabrohealaglYhIsle	1554	
Db	6120	GCGCGCTTGAAACACTGCC-----	GCTCGCGACAGG	6091
QY	1554	uProArgrglYserProValThrMetArgrgluProThrProArg-LeugIngluIsertL	1574	
Db	6090	CCCCCGCCGATCGCGCGCA-----CCACCAACCATCATCT-----	6056	
QY	1574	euserSerSerLYalaserGlInsrArgrglYsleuthrSerThrProArgrglu-----	1591	
Db	6055	-----ACCGGCCCCAGGACATCACCGGCATCCCGCATCCCGTACGACCCCA	6013	
QY	1592	-----lalaalyserProhIsertThrValProgluInhIsrProhIsrProI	1608	
Db	6012	TGCGGACCCACTCTCCACCAACGTCCAC--ACCAAGACCAACACACCCGACCCCA	5956	
QY	1608	leserProTYrglUhlIsleuLeuArgrglYalSerlYal-----Aspleut	1624	
Db	5955	TCCGGGAACACCCGCTCCCGGCGATGACGCAACCCATCCCGGATCCCGCAACTCC	5896	
QY	1624	YrArgrSerThIsleProleuAlaheArProThrSerleProArg-----	1639	
Db	5895	CCCGAACCACCCACCCGACCCGACA-CACCGGACCCACCAACACCCGCGCGCACCCG	5837	
QY	1640	-----glylleProleuAspAlaAlaAlaTYrT	1650	
Db	5836	ACCGAATCCCAACGACTCCACACCCCGGAACTCCCGCAACCGCTCATCA-----	5781	
QY	1650	YrLeuProArghIsleuAlaProArhArProThrTYrPro-HIsleuTYrProProTYrleu	1668	
Db	5780	-----ACAGGATGATMAACCATGCGACATCCCAATCCCAACCCCGGCACTC	5738	
QY	1670	-----lleArglYrTYrProArphrAlaAlaIsleuAlaAsrArgrglInthr	1684	
Db	5737	CACCCCATCCCCCATCAGCGAGGCA-CACCGCACTCCAGA-----	5697	
QY	1685	llelleAsrArpTYrleThrSerGlInlMetHIsIsArThrAlaThrAlaMeAla	1704	
Db	5686	-----CACCAATATCCCGGACACACCAACGACGCG	5664	
QY	1705	GlArgrAlaAsrMetLeuArgrglYleuserProArgrgluserSerleuAlaIsleuAsnTYr	1724	
Db	5663	GCCCATCTCA-----	5655	
QY	1725	AlaAlaGlYrProArgrglYllelleAsrleuserGlInValProhIsleuProValleuVal	1744	
Db	5654	-----CCGCGCAACGACACCCACT	5634	
QY	1745	ProProThrProglYrThrProAlaThrAlaMeArArgrleuAlaTYrleuProThrAla	1764	
Db	5633	CCCCCAACCTTCACAAACCCCGCGACACCGACTCCGAGCCCGCCACGACACATCC	5574	
QY	1765	ProGlInrProheserSerArghIserserSerProleuserProglYglYProThrHIs	1784	
Db	5573	CCCCACCCACGGGCAACCCACCATCAACCCACCGCGGCACCA-----CAACCGCA	5520	
QY	1785	leuthrlyrProThrInThrSerSerSerArgrgluArgrArgrArgrArgrArgrgluArgr	1800	
Db	5519	CCGCATCACCCCAACACAAACCCCGCGACATACGACGACGCAACTCCCGCCAGAT	5466	
QY	1805	AsrArgrArgrArgrgluArgrgluYserlleleuthrSerThrInThrValGlInhIsAla	1824	
Db	5459	GAC-----CCAAACACACGACACTCCACACCC	5433	

QY 1825 ProIleTPrArgProGlyThr-----GluInSerSer 1835
DB 5429 GAGCTTCACACGCCGGACACGCCACTCCAACGCCAAACGAGCGAGGTGAGCAAACT 5370
QY 1836 GlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgProAlaSerHisSer 1855
DB 5369 CCGTCCGACCCCAACACCCACCAACAGAACCGGACCAAGAACCAACCAACCC---CAGCACCG 5313
QY 1856 HisAlaHisGlnHisSerProIleSerProArgThrGlnAlaGlnGlnArgPro 1875
DB 5312 CACCGGACACCGGATCAGCCCGGACCCGCGTCA-----CGTCACTGCG 5265
QY 1876 SerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysPro 1895
DB 5264 CGAACATCATACATCCGCA-----ACGACCGACATCCACC-TCCCC 5224
QY 1896 ThrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaThrPheProPro 1915
DB 5223 ACCATGCAACACCTCATCCAGACCTCCGAAACACCCACCCCGCGATCAACCCA 5164
QY 1916 -----AlaThrHisCys-----ProLeuGlyGlyThrLeuAspGlyValTyr 1929
DB 5163 CGCCCATACCAACCACTCCGCACTCCGACCTGACCTGTACACAC--- 5119
QY 1930 ProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGlu 1949
DB 5118 CCCACC-----CCACCGGACACGCGCGCGGAC 5089
QY 1950 ArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProAlaArgSerGlyLeu 1969
DB 5088 CGCACACCCCGACACAC----- 5071
QY 1970 GluProAlaSerSerProSerLysGly-----SerGluProArg---Pro-LeuValPr 1986
DB 5070 CTCCTCGGACAAACCAACCAAGACGACGACGAGCTCCCGACCCACACACGCGCAC 5011
QY 1986 ArgProAlaSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHis 2006
DB 5010 ACCGCAACGGGTCTGAAAGAGTCCGGTGCAGAACCGCAATAG-----CCACATCC 4957
QY 2006 AlaSerProAspProPro-----AlaProProAlaSerAlaSer 2019
DB 4956 CGGAGATCCGACACCCGCTCCAGTACTACCCAAACCGACCGCTCCGCGCGCAC 4897
QY 2019 ArgProHisArgGlyLysThr--GlnSerLysProPheSerIleGlnGluLeuGluLeu 2039
DB 4896 GCGCCCTCCGACCGCGGACCAACACGCG-----CGAC 4861
QY 2039 ArgSerLeuGlyTyrHisGlySerSerTyr---SerProGluGlyValGluProValSerP 2058
DB 4860 GAGGACACCGCGTCCGCGGTACCGCGGTACGAGACGACGCGCGTCAATGTCCTGACC 4801
QY 2058 roVal-----SerSerProSerLeuThr 2066
DB 4800 CACTCCGCGGCTCTCGATGATCAATGCGCGTTCCTCCACTGATCCCGAACGACGAC 4741
QY 2066 HisAspLys-----GlyLeuProLysHisLeuGlnGluLeu 2078
DB 4740 ACAGCGCGCGCGCGCGCGCGCTCCGACCGCGGCGCGCTTCGCGACGAG 4681
QY 2078 ArgLysSerHisLeuGlnGlyGluLeuArgPro-----LysGlnProGlyProValLys- 2095
DB 4680 GC-----GACGGTGCAGGACCAATCGACGTCGTCGGCGCGGTCCGCG 4633
QY 2096 -----LeuGlyGlyGluAlaAlaHisLeuProHisLeu 2107
DB 4632 TGAAGTCCGCGCGACGCGTGCCTCA--GGGCAAGCAGCATTTGACACACCGG 4574
QY 2107 ArgProLeuProGluSerGlnProSerSerSerProLeuGlnThrAlaProGlyVal 2127
DB 4573 GCGCC-----CCGCGCGCGG 4559
QY 2127 ArgLysHisGlnArgValAlaThrLeuAlaGlnHisSerGluValIleThrGlnAsp 2147

DB 4558 CTGGGTGTGACCGACGTTGACTTCAGCGACCGGACCC----- 4521
QY 2147 YrThrArgHisHisPProGlnGlnLeuSerLarProLeuProAlaProLeuTyrSerPheP 2167
DB 4520 --ACAAGGGGTGCGCA-----CCCGCTCACCGCGGTGCGG 4484
QY 2167 roGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProP 2187
DB 4483 CAGCAGCGCGCTCGCTCCATCGAT---CGCCAGCGCGCTCCCGTACCGTCCGCTC 4427
QY 2187 roProAspHisGlyAlaProAlaArgGly-----S 2197
DB 4426 CACCGGTCCAGCTACCCGCGCGGACGCGCGTCCGACGAGGCGAGCTCGATGACGCC 4367
QY 2197 ArgProHisSerGlyGlyGlyLysArgSer----- 2206
DB 4366 CTCCTCGCGGACGGGTGGAGCGGTCAAGCGCTTGCTGCAACCTGCTTGACCGC 4307
QY 2207 -----ProGluProAsnLysThrSerValLeuGlyGlyAlaAspGlyIleGluProV 2225
DB 4306 ACTGCCCCGACACCG----- 4290
QY 2225 alSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuL 2245
DB 4289 -----CGAGACCGCGGTCTCCGACGCGCGGCGGTGAGAGCGCTCCAGTCCAGTAC 4235
QY 2245 euTyrArgAspArgGlyGlnGlnThrGluProSerArgMetGlySerLysSerProGlyVal 2265
DB 4234 G-GCGACACCTTGAGACAGGCGGTACG-----TGCGATCGCGGAGAGACG 4188
QY 2265 hrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLys 2285
DB 4187 GCATGACGCGCGCTGCGGGGCGAGCCCGCGCTCCCGCAAGACTCAGC----- 4138
QY 2285 exLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGlyTyr 2305
DB 4137 -----AACATGCGCGGCGCGCATCA 4116
QY 2305 snIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuM 2325
DB 4115 CCGTGGCGCGCGCGCGGACGACCGCGCTCGCGCG-----CGACGCGCT 4065
QY 2325 exThrTyrArgSerGlnAlaVal-----GlnGlnHisAlaSerThrAsnMetGlyLeu- 2342
DB 4064 GCACGGCGACGTGACGCGCCACCAAGGACGAGAGACGCGCGTGCAGCGTACGCGCG 4005
QY 2343 -----GluAlaIleIleArgLysAlaLeuM 2351
DB 4004 GTCCGCGGTCGAGAGGTAGCGATCCGCGCGGAGGACGCTCGCGGTGTCGCG 3945
QY 2351 exGlyLysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProL 2371
DB 3944 TGAGCAGTACCTCTGTATCCGTCCGCTCCCTCGTGAAGCTCGCGCGCTACTCCGCTG 3885
QY 2371 euAsnAlaSerLysLeuProAlaAlaMetPro-----IleThrAla 2386
DB 3884 CGGTGCGCGGACGAAACATCCGCGCTCCGCGCGCGCACTACTGCGGAGACCTCCG 3825
QY 2386 HisAspGlyArgSerAspHis-ThrLeu-----ThrSerProGlyGly 2399
DB 3824 CGTCTCGAGGCGCTCCACGACGCTTCAGACGACGAGCGCTGCGGATCATGCGCG 3765
QY 2400 GlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaPro 2419
DB 3764 TGCGCTCGCGCGGATTC---CGAAGACTCGCGTCACTCGCGCGG----- 3714
QY 2420 GlyLeuAlaSerGlyAspArgProPro-----SerValSerSerValHisSerGlu 2436
DB 3713 -----AGTGCAGAACCGCGCTCCGACCATAGTCTTGGCGGCGCTCCGCGCTCG 3660
QY 2437 GlyAspCysAsnArgArgThrProLeuThrAsnArgValITrpGluAspArgProSerSer 2456

Db 3659 GGTGGTACAGCCCGCTCAGAGTCCAG -CCGGGCTCGGTGGGGAAGACAGATGGCTCG 3601

Qy 2457 ALaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMet 2476
3600 GCG-----CCGAGCTCCAGACAGCTCCAGAGAGCGGCGCG 3655

Db 2477 ALaSerProPro-----ProGlyLeuProAlaGlySerGlyPro 2490
3504 GAGTCGACGCCGCCCGGGAGCCGCGACGCCGACACCCACGATGGCGATGGGTCGTGCTCG 3505

Qy 2491 LeuAlaGlyProHisAlaTyrAsp-----GluGluProLysPro-- 2504
3504 TCCGGGCTGCGGCCCGGCGGTCCGACACCCCTACCGCGCGCAGAGACGATCCGGGCG 3445

Db 2505 -----LeuLeuGlySerGlnTyrGluThrLeuSerAspSerGlu 2517
3444 CCGCCGCGCTCGAAGCTTCTGTGGCGCCGTCGAGAGACGAGCCCGAGGGGAG 3391

Search completed: November 12, 2005, 14:54:06
Job time : 41449 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_p2n model

Run on: November 11, 2005, 09:34:07 ; Search time 35711 Seconds
(without alignments)
3415.244 Million cell updates/sec

Title: US-09-522-753-5
Perfect score: 13215
Sequence: 1 MSGSTOLVAQTWRATEPRYP.....WDEEPKPLCSQYETLSDSE 2517

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xdp
-Q=cg2_1/USFTO.epool_p/US09522753/runat_02112005.173640.20793/app.query.fasta.1.2695
-DB=GenBml -QFMT=faetap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi -LIST=45
-DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=prc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09522753 @CGN 1.1 14601 @runat 02112005.173640.20793 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenBml.*
1: gb_ba.*
2: gb_hg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	13215	100.0	8561	9	AF113003	AF113003 Homo sapi
2	12978	98.2	8686	9	AF125672	AF125672 Homo sapi
3	12922	97.8	9053	6	AR447713	AR447713 Sequence
4	11067	83.7	8388	10	AF113001	AF113001 Mus muscu

5	10832.5	82.0	8544	10	AF125671	AF125671 Mus muscu
6	9894	74.9	7465	10	AF113002	AF113002 Mus muscu
7	9776.5	74.0	6339	6	C0722208	C0722208 Sequence
8	7885	59.7	5989	9	HSU37146	U37146 Human gllen
9	4612	34.9	2930	9	S83390	S83390 T3 receptor
10	4194	31.7	7949	9	AB028970	AB028970 Homo sapi
11	4187	31.7	7780	10	MM035312	U35312 Mus musculus
12	4147.5	31.4	7940	6	C0847916	C0847916 Sequence
13	4147.5	31.4	7940	6	AX578069	AX578069 Sequence
14	4147.5	31.4	7940	9	AF044209	AF044209 Homo sapi
15	4043	30.6	8018	9	AF087856	AF087856 Homo sapi
16	4009.5	30.3	8959	9	AF495886	AF495886 Xenopus 1
17	3736.5	28.3	191100	10	AC139377	AC139377 Mus muscu
18	3479.5	26.3	254449	2	AC097560	AC097560 Rattus no
19	3445	26.1	218129	2	AC136560	AC136560 Rattus no
20	3344.5	25.3	219339	2	AC121005	AC121005 Rattus no
21	3178	24.0	2842	9	BC004326	BC004326 Homo sapi
22	3012	22.8	2964	10	BC047524	BC047524 Mus muscu
23	2910.5	22.0	6300	6	C0724013	C0724013 Sequence
24	2891.5	21.9	205283	9	AC073916	AC073916 Homo sapi
25	2187	16.5	3120	9	AF303586	AF303586 Homo sapi
26	2155	16.3	1937	5	BC054296	BC054296 Xenopus 1
27	2147	16.2	1836	5	BC082706	BC082706 Xenopus 1
28	2091.5	15.8	6328	10	AB093281	AB093281 Mus muscu
29	1900	14.4	6541	9	AB019524	AB019524 Homo sapi
30	1839	13.9	4285	10	MUSRIP13G	L78294 Mus musculu
31	1687	12.8	1891	9	BC050594	BC050594 Homo sapi
32	1663.5	12.6	1850	9	BC058511	BC058511 Homo sapi
33	1645.5	12.5	1808	9	BC068996	BC068996 Homo sapi
34	1633	12.4	1741	9	BC056862	BC056862 Homo sapi
35	1628.5	12.3	3025	5	BC049302	BC049302 Danio rer
36	1611	12.2	2145	5	BC073704	BC073704 Xenopus 1
37	1530.5	11.6	1857	5	BC073253	BC073253 Xenopus 1
38	1396	10.6	161970	2	AC027706	AC027706 Homo sapi
39	1391	10.5	2914	6	C0850622	C0850622 Sequence
40	1391	10.5	2914	9	AK127788	AK127788 Homo sapi
41	1350.5	10.2	1152	5	AY498876	AY498876 Xenopus 1
42	1172.5	8.9	3997	9	AF303585	AF303585 Homo sapi
43	1162	8.8	133947	5	AL590153	AL590153 Zebrafish
44	1113	8.4	1650	6	AX677743	AX677743 Sequence
45	1099	8.3	11236	3	AF175223	AF175223 Drosophila

ALIGNMENTS

RESULT 1
LOCUS AF113003 8561 bp mRNA linear PRI 20-MAR-1999
DEFINITION Homo sapiens silencing mediator of retinoic acid and thyroid hormone receptor alpha mRNA, complete cds.
ACCESSION AF113003
VERSION AF113003.1 GI:4454551
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 8561)
ORDentlich,P., Xie,W., Genin,A., Spinner,N.B. and Evans,R.M.
TITLE Unique forms of human and mouse nuclear receptor corepressor SMRT
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (6), 2639-2644 (1999)
MEDLINE 99178941
PUBMED 10077563

A0220946

REFERENCE
AUTHORS Downes,M.R., ORDentlich,P. and Evans,R.M.
TITLE Direct Submision
JOURNAL Submitted (11-DEC-1998) Gene Expression Laboratory, The Salk
Institute for Biological Studies, 10010 North Torrey Pines Road, La
Jolla, CA 92037, USA
FEATURES
source
1..8561
/organism="Homo sapiens"

CDS

/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12q24"
/feature_type="pituitary"
2..7555
/function="transcriptional co-repressor"
/note="hSMRT alpha; longer isoform than previously reported"
/codon_start=1
/product="silencing mediator of retinoic acid and thyroid hormone receptor alpha"
/protein_id="AAD20946.1"
/db_xref="GI:4454552"
/translation="MSGSTQLVAQTRWATEPRYPHSLSYPVQIARTHTDGLLEYOH
HSDRYASHLSPGSIIOFORRRPILQPSGPNERSOELHRLPSHLYDELKSEMEF
TESKPRLEILPDILRPSPLATQSPASEDLTKDRSLTKLEHSPSPSPHTDBEL
ELVPRRLSKEBILQNMNDYREITWEEQISLKAKKOQDEBBAKAPPEKPEVSP
ISKRSILVQIITYDENRKAEMARILBGLQFOVELPLYNQPSDTRQYHENTKINAM
RKKLILYFRRHARQWQKQCORYDQMEALEKKVERIENNPRRAESVRYEYE
KQPEIRKQREIQRMSVQVQSGLSMSAARSEHEVEIIDNLSQENLEKQMKOL
AVIPMLYADQOQRIKFIMNGMLADPMKYKQRYVMNMSQEKETPFKKMKQKN
FGLASFLEKTVABECVLYYLTKKNKNTKSLVRSYRRGRSQOQOQOQOQOQOQ
QOPMRSSQEBKDEKKEKKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKE
KORRANSSQRRKGRITRSMANBANSEAITTPOQASLASMELNESRTEBMEYAK
KCLLEHGRWMSAIARWGSKTVSQCNFYFNTKKQNLDEIILQHKLMKEKRNARX
KKKAPAASEBAAPRVVEDEMEASVGENEWEVEEALHAGNEVRESCGPA
TVNNSDTEGSIPTSPTEAKDQNGPKPATGADGPPGPPPTPRRTSRPIETPT
ASEATGAPTPPAPPSAPAPPVYKKEKEEETAAAPVBEGBEOKPPAAEELAVTG
KAEPRKSECTSEAEGBGAKGDABAABATAGALKKKEKGGSGAATTAKSGADOD
SSSATSADVEDEAGBKNDLSPRESLTPTGDPANASPOKTLDKOLKORAA
IPIPIQVTKNEPPREDAAPTKAPAPAPPPQNIQPESSDAPQOPSSPRKSSPAPA
DEAFPAEAKLPGDPFCWTSGLPFVPREVLKASPHADBSAFSAPRGHPPLGL
HDTARVLRPPTISNPPLTISAKPSVLEROIQAIISQMSVOLPYSEHAKAVG
PYTMGILPMDPKKLAPSGSGQEOQLSPGQAGPESLIVPTAOEASVLRGALSVP
GGSITKGTSTKVPDSALTYRGSITNGRPADVLYKGTITRITIGBSPERLDGRDS
LKRGHVTEGKGHVLISYEGMSVTQCSKEDKSSGPHETNAPKRTITDMEGRGR
AASSASIEGLMGRALPERHSPHNLKEQHNIRGSIIOGIPRSVKAQEDYLREAKLL
KEEGTPPPPPRDLTEAVKTOALGPKLPAHEGLIYAVKEGSRHBIPEBELHT
PELAPARPLKSGSITQGTPLKYDTGATSGTSGKHDVRSILGSPRTPPVHLDVMA
DRALEKACVESIKSRPGTASSGGSIARGAPVYVBEIGKSGSLTYEDGDAPAG
HLPKPSPTMREPTPLQSGLSSSKASQDRKLTTPREIASKPSHTVEHNHPISP
YELHLKGVSGVDLYRSHIPLADPTSIIPGILPDAADAAVYMLRNLNAPHTYLPY
LIRGLDGTALENRQTIINDYITISQOMHNITATAMQADMLKSLPRSSIALNVA
GREGIIDLQVPHLPVLPPTGCPATNADPLAYLPTAOPFSRSSSSPLSGGPTH
LKPPTTSSEREDRDRDREREKILISTTVBAPIPRCTPSSSSSSGSGSG
TGGSSSRPASHAHQHSPISPRTODAIQOBPSVILNMGMIITAVPNSPTVRS
TSSSEVPRAATTPPATHCPLGSTDGVTPLMEPVLILKREKAVRAPERPADTSHA
FLAKPARSGLEBPASSPKSGSEPRPLVPVSGHATTAKTPAKNLAHNASPPAPPA
SASDPHREKTOGKSPSIOELSLASLGHSSYSPEGVBPVSPSSPLTHDGKPLML
BEIDKSHLEGEIARQKPGPVKLGGEAHLPHLPLPESQPSSSPLQITAPAGYKHQV
VTLAHOHISEVITODYTRHNPQOLASPLAPLVSFGASCPVLDLRSPADVLYLPD
GAPAGSPHSBGGKRSPEHNKTSVIGGGGDIIEPVSPPGMEPGHRSRSPVLYLD
GSGTSESRMGSKSPGNTSOPPARFSGLTSSNMGMYSKQEIINKLNTNRNRENTI
SQGTETFNMPAITGTGLMTYRSQAVQDEASTWGLEAIIKRLMKSLQKMSPEPLS
AANAFNLNASILPAMPITTAADGRSDHLLTSPGGGKKAKVRSRPSRKAQBPAPLA
SGDRPVSVSVEGDGNCNRTPLTNWVWDPRSAGSTFPFVPLIMRLQAGVMAASP
PGGLPAGSGPLAGPHHAWDEBPKPLCSQYETLSDSE"

ORIGIN

Alignment Scores:

Pred. No.: 1,286-258 Length: 8561
Score: 13215.00 Matches: 2517
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-522-753-5 (1-2517) x AF113003 (1-8561)

QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTyrArgAlaThrGlnProArgTyrPro 20
DB 2 ATGTGGAGCTCCACACAGCTTGTGGACACAGCTGGAGGGCCACTGAGCCCGCTCACTCCG 61

QY 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
DB 62 CCCACAGGCTTCTCTACCCAGTGCAGATCGCCCGACGACAGGACCTCGGGCTCTCG 121
QY 41 GlnTyrGlnHisHisSerThrArgAspTyrIleSerHisIleLeuSerProGlySerIleIleGln 60
DB 122 GAGTACACAGACACTCCCGGACATATGCTCCACCTGTGCGCGGGCTCTCATCATCCAG 181
QY 61 ProGlnArgArgProSerLeuLeuSerGlnProGlyAsnGlnArgSerGln 80
DB 182 CCCACGGCGGAGGCGCTCTCGTGTGATTCAGGTCGCGGGAAATGAAGGTCGAC 241
QY 81 GlnLeuHisIleLeuArgProGlnSerHisSerTyrLeuProGlnLeuGlyIleSerGlnIle 100
DB 242 GAGCTCCACCTCGCGGACAGTCCCATCATACCTGCCCCGAGCTGGGAAAGTCAGAGATG 301
QY 101 GlnPheIleGlnSerTyrAspProArgLeuGlnIleLeuProAspProLeuLeuArgPro 120
DB 302 GAGTTCATTGAAAGCAAGCGCTCGGTGAGCTGCTGACCCCTGCTGCAACCG 361
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGlnAspLeuThrIleAspArgSer 140
DB 362 TCACCCCTGCTGCGCACGCGGCGAGCTGGGGATCGAAGACTCACCAAGACCGTAC 421
QY 141 LeuThrGlyIleLeuGlnProValSerProProSerProHisThrAspProGlnLeu 160
DB 422 CTGACGGGCAACCTGGAACGGGTGCTCCCGCCAGCGCCCGACACTGACCTGAGCTG 481
QY 161 GlnLeuValAlaProProArgLeuSerTyrGlnIleGlnIleGlnAsnMetAspArgValAsp 180
DB 482 GAGCTGGTGCCTCCACGCGGTGTCCAAAGAGCTGATCCAGAACATGGAACGCGTGAC 541
QY 181 ArgGlnIleThrMetValGlnGlnGlnIleSerTyrIleLeuIleLeuSerGlnGlnIle 200
DB 542 CAGAGATCAACATGTTAAGCAGACGATCTTAAGCTGAAGAAAGCAGCAACGCTG 601
QY 201 GlnGlnGlnAlaAlaIleAspProGlnProGlnIleProValSerProProProlIleGln 220
DB 602 GAGGAGGAGGCTGCCAAGCGCGCGGAGCTGGAAGCGCGGTGTACCGCGCCGCACTGAG 661
QY 221 SerIleHisArgSerLeuValGlnIleIleTyrAspGlnAsnArgIleValIleGlnAla 240
DB 662 TCGAAGCACCGGACCTGTGAGATCATCTACGACGAACCGGAAGAGCTGAAGCT 721
QY 241 AlaHisArgIleIleGlnGlnIleLeuGlyProGlnValGlnLeuProLeuTyrTrasnGlnPro 260
DB 722 GACATCGGATTTCTGGAAGGCTGGGGCCCGAGGTGAGACTGCGCTGTACAAACGAGCC 781
QY 261 SerAspThrArgGlnTyrHisGlnAsnIleIleValIleAsnGlnAlaMetArgIleValLeu 280
DB 782 TCCGACACCCCGGACAGTATCATGAACAATTAACACAGCGCATTCGGGAAGAGCTA 841
QY 281 IleLeuTyrPheIleArgArgHisHisAlaArgIleGlnTyrGlnIleIleAspPheCysGln 300
DB 842 ATCTTGTACTTCAAGAGGGAHAATACGCTCGAAGAACATGGAAGCAAGATTTCTCCAG 901
QY 301 ArgTyrAspGlnLeuMetGlnAlaLeuGlnIleValIleValIleGlnIleValIleGln 320
DB 902 CGCTATGACACAGCTCATGAGAGGCTTTGGAAGAAAGGTGAGCGCATGAAAAACACCGG 961
QY 321 ArgArgArgAlaIleGlnIleSerIleValIleArgGlnTyrTyrGlnIleGlnIlePheProGlnIle 340
DB 962 CCGCGGCGGCGCAAGAGAGCAAGGTCGCGGAGTCTACGAAAGAAAGATTTCTCGAGATC 1021
QY 341 ArgIleGlnArgIleLeuGlnGlnIleValArgMetGlnIleArgValIleGlnIleArgIleSerGly 360
DB 1022 CCGAAGCAGCGGAGCTGAGAGGCGCATGACAGAGGAGGTGGCCACGCGGCGCATGGG 1081
QY 361 LeuSerMetSerAlaIleArgSerGlnIleGlnIleValSerGlnIleIleAspGlyLeuSer 380
DB 1082 CTGTCTCATGTGGCGCGCGCGGACGAGCAGCAGAGGTCTCAGAGATCATCATGAGCTCTCA 1141

QY 381 GUGNGUUAUNLEUGULYVSGIMETARGGULEUAVALIIEPROMETLEUTYR 400
DB 1142 GAGCAGAGAGAACTGGAGAGACATGCGCAGCTGGCGTGATCCGCCATCTGTAC 1201
QY 401 ASPALAAPGNGINAGYIIELVPHGELLASMETANGIYLAUMETALAAPROMET 420
DB 1202 GACCTTACACGACGCCATCAAGTTCAATCAATGAACGGGCTTATGGCCGACCATG 1261
QY 421 LYEVALLTYRILYSAAPATGGINVALIMETASMETTPSERGIUGINGULYVSGIUTHRPHE 440
DB 1262 AAGGTGACAAAGACCCCGAGCTCATGACATGTGAGTGGACGAGAGAGAACCTTC 1321
QY 441 ARGGULYVPHMETGNGIHISPROLYASBNPHEGLYEULIIEALASERPHLEUGIYARG 460
DB 1322 CGGAGAAAGTTATGACAGCATCCCAAGAACTTGGCTGATCGCATCTCTGAGAGG 1381
QY 461 LYETHRALIAGIUCYVVALILEUTYRTTYRILEUTHRYLYBSANGIUAERTYLYS 480
DB 1382 AAGACATGGCTGAGTGGCTCTTATTACTGACTGAAGAAATGAACTATPAG 1441
QY 481 SERLEUVALIARGATSERTYRARGARGGILYVSSERGINGINGINGIN 500
DB 1442 AGCTGTGAGACGAGACTATCGCGCGCGCAAGCCAGCAGCAACAGCAGCAG 1501
QY 501 GINGINGINGINGINGINGINGINGINPROMETPROARGSERSEGINUGIULYS 520
DB 1502 CACGACAGCAGCAGCAGCAGCAGCAGCCATGCCCGCAGCAGCAGAGAGAGAA 1561
QY 521 AARGGULYVSGIULYVSGIUALAGIULYVSGIUGIULYVPROGIUVALIUIASN 540
DB 1562 GATGAG 1621
QY 541 AARGGULYVSGIULYVSGIULYVTHRAPASPTHSERGIYGIUAAPASAPGJU 560
DB 1622 GACAGAGAGAACTCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1681
QY 561 LYVSGIUALAVALIASERLYSGIYARGLYVTHRALIASNSEGINGIYATGATLYVSGIY 580
DB 1682 AAGAGAGCTGTGGCTCCAAAGGCCGCAAACTGCCAAAGCCAGGAGAGAGAGAGAG 1741
QY 581 ARGIIIEHRARGSERMETALASNGLUALASNSEGIUGIUALIIETHRPROGIN 600
DB 1742 CGCATCACCCGCTCAATGGCTAATGAGGCCAACAGCAGAGAGGCAATCACCCCCACAG 1801
QY 601 SERIAGIULEUAIASERMETGIULEUANGIUSERSERATGTPTHTGIUGIULMET 620
DB 1802 AGCGCCAGCTGGCTCTCATGAGACTGAAATGAGATTCGCTGGACAGAGAGAGATG 1861
QY 621 GIUTHRALALYVBSGIYLEULEUGIUIHISGLIYARGANTTPSERALIALIARGMET 640
DB 1862 GAAGAGCCCAAGAAAGGCTCTCTGGAAACAGCGCGCAACTGCTGCGCATCGCCGGATG 1921
QY 641 VALIGLYSERLYSTHRVALISERGINCYVLYASBNPHEPTYRPHENITYLYVLYARGIN 660
DB 1922 GTGGGCTCCCAAGACTGTGTGCGAGTGAAGAACTTCACTCACTCAAGAGAGGCGAG 1981
QY 661 ASNLEUASPGIUILELEUGINGIINHISLYVLEULYSEWETGIULYVSGIUAARGAANALARG 680
DB 1982 AACCTGATGATGATCTTGACAGCAGCAACAGCTGAAGTGAAGAGAGAGAGAGAGAGAG 2041
QY 681 ARGLYVLYVLYVLYVLYVLYVLYVLYVLYVLYVLYVLYVLYVLYVLYVLYVLYVLYVLYV 700
DB 2042 AGGAG 2101
QY 701 GIUAAPGIUGIUMETGIUALASERGIYVALISERGIYASNGIUGIUGIUMETVALIGIUGI 720
DB 2102 GAGAGATGAG 2161
QY 721 ALAGIUALIAREUHIISALASERGIYASNGIULVALI PROARGIYGIUCYVSSERGIYPROALA 740
DB 2162 GGTGAAGCCCTTACATGCTCTTGGAATGAGAGTGCAGAGAGAGAGAGAGAGAGAGAGAG 2221
QY 741 THRVALIASNANSESERAPTHRGIUSERILEPROSERPROHISTHRGIUALIALALYS 760

DB 2222 ACTGTCAACAAACAGCTGACACACGAGAGCATCCCTCTCTCACATGAGGCGCGCAAG 2281
QY 761 ASPTHRGIYGINANGIYPROLYVPROPROALATHREUGIYALIASPGLIYPROPRO 780
DB 2282 GACACAG 2341
QY 781 G1YPROPROTHRPROPROARGATGTHRSERARGIAPROIIIEGIUPROTHRPROALASER 800
DB 2342 GGCCCAACCAACCCACAGAGAGACATCCCGGGCCCCCATTTAGGCCACCCCGGCTCT 2401
QY 801 GIUALATHRGIYALAPROTHRPROPROPROALAPROPROSERPROSERALAPROPRO 820
DB 2402 GAAGCCACCGGAGGCTCTGAGCCGCCACAGACCCCATGCGCTCTGCACTCTCTCT 2461
QY 821 VALIVALIYVSGIUGIULYVSGIUGIUGIUTHRALIALIAPROPROVALIGIUGI 840
DB 2462 GTGGTCCCAAG 2521
QY 841 GIUGIUGIULYVPROPROALIALAGIUGIULEUALIIVALIAPTHRGIYLYVSAAGIUGI 860
DB 2522 GAGGAG 2581
QY 861 PROVALIYSSERGIUCYSTHRGIUGIUALAGIUGIULYVPROALALYSGIYLYVSAAPALA 880
DB 2582 CCGGTCAAGAGCAGATGTCAG 2641
QY 881 GIUALIAGIUALATHRALIAGIUGIYALILEULYVALIAGIULYVBSGIUGIYVSGIY 900
DB 2642 GAGGCGCTGAGGCAAGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2701
QY 901 G1YARGIATHRTHRALIYVSSERGIYALAPROGINASPSERAPSERSERIATHR 920
DB 2702 GCGAGGCGCCACACTGCAAGAGCTCGAGGCGCCCCCGAGAGACAGCACTCAAGTCTAAC 2761
QY 921 CYSESERIALAAPGIUVALIAPGIUVALIAGIUGIYVGIYVAPLYASBNATGLEULEUSER 940
DB 2762 TGCAGTCAACAGAGGTGATGAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2821
QY 941 ARGPROSERILEUTHRPROTHRGIYVAPPROARGIALASNALASERPROGINLYVPRO 960
DB 2822 AGGCCAGCTCTCTCACCCGAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2881
QY 961 LEUASPLEULYSGIULEULYSGIINARGALIALIAPROTHRYLEUVALIATHRYLYS 980
DB 2882 CTGAGCTGAAGAGAGCTGAAG 2941
QY 981 VALIAGIUBROPROARGIUAAPALIALIAPROTHRYLEUVALIAPROPROALAPROPRO 1000
DB 2942 GTCATGAG 3001
QY 1001 PROPROGINASPLEULYSGIINARGALIALIAPROTHRYLEUVALIATHRYLYS 1020
DB 3002 CCAAGCCCAAACTCTGAG 3061
QY 1021 G1YLYVSSERARGSERPROALAPROPROALIASPGLYVLYVLYVLYVLYVLYVLYVLYVLYV 1040
DB 3062 GGCAG 3121
QY 1041 LYVLEUPROGIYVAPPROPROCYSTPTTHRSERGIYLEUPROPHNEPROVALIAPROPRO 1060
DB 3122 AAGCTGCTGGAG 3181
QY 1061 GIUALIILIELYVALIASERPROHISALAPROASPPROSERALIASERTYRVALIAPROPRO 1080
DB 3182 GAGGTATCAAG 3241
QY 1081 G1YHISPROLEUPROLEUGIYLEUHIISAPTHRVALIARGPROVALIENUPROARGPROPRO 1100
DB 3242 GGTACACCACTGCGCTTGAGGCTCTCATGACACTGCGCGGCGCTCTGCGCGGCCACCC 3301
QY 1101 THRILESERASNPROPROLEUUIESERSERIALIYHISAPROSERVALILEUGIYARG 1120

Db	3302	ACCAATCTCCAAACCCGCTCCCTCTATCTCTCTCTGCAAGCACCCCAAGGTCTCTGAAGAG	3361
Qy	1121	GINIIIEGIALIAIIIESerGInGIyMeSeSerValGInIuEHISValProTySeSerGIuHIS	1140
Db	3362	CAAAATAGGTCATCTCCCAAGAAATGTCGGTCCAGCTCCACGTCCCTCATCTCAAGACAT	3421
Qy	1141	AlAlYsAlAProValGIYProValThrMetGIYLeuProLeuProMetAAsPProLYs	1160
Db	3422	GCCAAAGGCCCGGTGGGCTGTTCACATGGGGCTGCCCTGGCCATGACCCCAAAAG	3481
Qy	1161	LeuAlAProPheSeSerGIYValIysGInGIuInLeuSeSerProArSGIYGIAlAgiYPro	1180
Db	3482	CTGGCACTCTTCAGCGAGGTAGAGAGACAGACTGTCTCCACCGGGCCAGGCTGGGCCA	3541
Qy	1181	ProGIuSeSerLeuGIYValProThrAlaGInGIuAlaSeSerValLeuArGIYThrAlaLeu	1200
Db	3542	CCGAGAGAGCTGGGGGTGGTCCCAACGCCAGAGGGGTCCGTGTGAGAGGACAGCTCTG	3601
Qy	1201	GIYSeSerValProGIYGIYSerIIEthryIsGIYIIEProSeSerThrArGIYProSeSerAAsP	1220
Db	3602	GGCTCAGTTCGGGGCGGAAGCATCCAAAGGCATTTCACAGACACGGGTGCTCCCTCGGAC	3661
Qy	1221	SeSerAlaIIEthrTyArGIYSerIIEthrhISGIYThrProAlAAsPValLeuTyLYs	1240
Db	3662	AGCGCACTACATACCGGGGCTTCATACCCACGGACGCCAGCTGACTGTCTGTACAG	3721
Qy	1241	GIYThrIIEthArGIIEIIEGIYIuAAsPSeSerProSeArGIYLeuAAsPArGIYArGIu	1260
Db	3722	GGCACCATCATCCAGGATCATGTGGCGAGGACGCCGACGTCCGTTGGACCGGGGCCGGAG	3781
Qy	1261	AAsPSeSerLeuProLYsGIYHISValIIEtyrGIuGIYLYsLYsGIYHISValLeuSeSerTy	1280
Db	3782	GACACCTGCCCCAAGGGGCACCTCATCTACGAAGGCAAGAGGGGCACAGTCTTGCTCAT	3841
Qy	1281	GIuGIYGIYMeSeSerValThrGInCySeSerLySGIYuAAsPArGIYArSeSerSeSerGIYPro	1300
Db	3842	GAGGGTGGCATGTCTGTGACCCAGTGCTCCAAAGGAGACGGAGAAAGAGCTCCAGAGACC	3901
Qy	1301	ProHISGIuThAlAlAProLYsArGIYThrTyArAAsPMeTmetGIuGIYArGIYArGIYArG	1320
Db	3902	CCCCATGAGACGGCCGCCCCCAAGGACCTATGACATGATGAGAGGGCCGGTGGGACGA	3961
Qy	1321	AlAlIIESeSerSeSerAlaSerIIEGIuGIYLeuMeTGIYArGIAlAlIIEProProGIuArGIHIS	1340
Db	3962	GCCATCTCTCTGACGACGATCGAAAGTCTCATGGGCTGGCCATCCCGCCGAGCGACAC	4021
Qy	1341	SeSerProHISHISLeuLYsGIuGInHISHISAlIEArGIYSerIIEthGInGIYIIEPro	1360
Db	4022	AGCCCCCACCACTTAAAGAGGACCAACAATCCCGGTGCTCATCAACAAGGATCCCT	4081
Qy	1361	ArGSerTyYrValGIuAlaGInGIuAAsPtyrLeuAlASeSGIYuAlAlAlyLeuLeuLYsArG	1380
Db	4082	CGGTCTTACGTGAGGACAGAGGACATACCTGCGTCCGAGGCCAAGCTCTTAAAGCGG	4141
Qy	1381	GIuGIYThrProProProProProSeSerArGAAsPLeuThrGIuAlATyryLYsThrGIn	1400
Db	4142	GAGGGCAGCGCTCCGCCCCACCGGCTTCACGGGACGTGACCGAAGGCTTACAAGCGCAG	4201
Qy	1401	AlAlEugLYProLeuLYsLeuLYsProAlAlHISGIuGIYLeuValAlAlThrValLYsGIu	1420
Db	4202	GCCCTGGGGCCCCCTAAGCTGAAGCGGGCCCATGAGGGCTGTGGTCCACGGTGAAGAG	4261
Qy	1421	AlaGIYArGSerIIEHISGIuIIEProArGIuGIuLeuArGIHISThrProGIuLeuPro	1440
Db	4262	GGGGGCGGCTCATCATGAGATCCCGGCGAGAGCTGGCGACACGCCCGAGCTGCC	4321
Qy	1441	LeuAlAProArGIProLeuLYsGIuLYserIIEthrGInGIYThrProLeuLYsTyArP	1460
Db	4322	CTGGCCCCCGCGCTCAAGGAGGGCTCATCAACGAGGACCCCGCTCAAGTACGAC	4381
Qy	1461	ThrGIYAlASeSerThrThrGIYSerLYsLYsHISAAsPValArGSerLeuIIEGIYSerPro	1480
Db	4382	ACCGGGCGCTCCACCACTGGCTCCAAAAGACGACGTAGCTCTCTCATGGCGGCCCC	4441

QY	1481	GLYATHRPHRPDPROYVAIHSPROLEUAPVALMETALAAPVALAARGVALALEUGLU	1500
DB	4442	GGCCGGAGGTTCCACCCGGTGCACCCGGTGAATGATGAGCCGACGGCCCTGGAA	4501
QY	1501	ARGVALCYRYTGLUGLUSERLEUYSERARGPROGLYTHRALASERSESGLYGLY	1520
DB	4502	CTGCGCTCTCAAGAGAGAGCCCTGAAGACCGGCCAAGAACCGCCACAGCTCGGGGGGC	4561
QY	1521	SERILEALARGGLYALAPROVALILEVALPROGLULEUGLYLYSPROARGLINSERPRO	1540
DB	4562	TCCATTGGCGCGCGCGCCCGCTCATTTGTGCTCAGCTGGGTAAAGCCGCGCAAGCCCC	4621
QY	1541	LEUTHTRYGLIUAAPHISGLYVALAPROPHETAGLYHISLEUPROACGGLYSEPROVAL	1560
DB	4622	CTGACCTATGAGAGACACAGGGGACACCTTTTCGGGACACCTCCACAGATGTCCCGCTG	4681
QY	1561	THMETARGIUPROTHPROARGLEUGLUGLYSERLEUSERSESERLYALASER	1580
DB	4682	ACCATGGGGAGCCCAACGCGCGCTCGAGGAGGGACGCTTTGTCCAGAAAGCATCC	4741
QY	1581	GLIDRAPGLYBLEUTHSERTHSERTHPROARGLUILEALYSSERPROHISERTHVAL	1600
DB	4742	CAGACCCAAAGCTGACCTGCACCCCTCTGTAGATCGCAAGTCCCGACAGACCGTG	4801
QY	1601	PROGLUHISHIPROHISERPROLSESPROTYGLIUNHISLEUEUARGLYVALSERGLY	1620
DB	4802	CCCGAGACCAACCCACACCCCATCTCGGCGCTATGAGACACCTGTTCGGGGGCTAGTGGC	4861
QY	1621	VALAPLEUITYRAARGSERHISILEPROLEUALAHESAPROTHRSERILEPROARGLY	1640
DB	4862	GTGAGCCGTATCGACAGCACATCCCCCTGGCCTTGACCCCACTCCATACCCCGGCGC	4921
QY	1641	ILEPROLEUAPRALAALAALATETRYTLEUPROARGHISLEUALAPROASNPOTHR	1660
DB	4922	ATCCCTCTGAGACGACGCGCTGCTTACTACTGCGCCGACACCTGGCCCCAACCCACCC	4981
QY	1661	TYRPROHISLEUITYRPROPROTYLLEUULEARGGLYTYRPROAPHTHALAALALEUGLU	1680
DB	4982	TACCGGCACTGTATCCACCCCTTACTCATCCGCGGCTAACCCGACACGCGCGCTGAG	5041
QY	1681	ASPARGGINTHRIELLEASAPADRYLLETHSERGINGLIMETHISIASHTHALA	1700
DB	5042	AACCGGACCATCATATATGACTACTACCTTCGACAGATGACACACACACACGCGC	5101
QY	1701	THRALAMETALAGLARGALAAAPMETLEUARGLYLEUSERPROARGLINSERSEUEN	1720
DB	5102	ACCGGCATGGGCCACGACGAGCTGATATGTGAGGGGCGCTCGCCCGGAGATCTCGCTG	5161
QY	1721	ALAIEUANTYRALAALAGLYPROARGLYLLEILLEAPLEUSERGINVALPROHISLEU	1740
DB	5162	GCACCTCACTACGCGCGCGGTCCCGAGGACATTCAGCTGTCCCAAGTCCACACTTG	5221
QY	1741	PROVALIEUVALPROPROTHPROGLYTHRPROALATHRALAMETAPARGLEUALATYR	1760
DB	5222	CTGTGCTCTGTCGCCCCCGACACACGACACCCCAACACCGCATAGAGACCGCTTGCTAC	5281
QY	1761	LEUPROTHRALAPROGINPROPHESERSEARGHISERSESERSESPROLEUSERPROGLY	1780
DB	5282	CTCCCAACCGGCGCCCAAGCCTTACGACGCGGCACAGCAAGCTCCCATCTCCCAAGGA	5341
QY	1781	GLYPROTHRIHISLEUTHLYSPROTHTRHTRHSESERSESGLYVALARGIUAARGAPARG	1800
DB	5342	GGTCCAAACACTTGACAAHACCAACCAACCAAGTCTGTCTCCAGCGGAGACGAGACCGG	5401
QY	1801	ASPARGIUAARGAPARGAPARGIUAARGIUNHYSERILEUTHSERTHTRHTRH	1820
DB	5402	GATGAGAGCGGGACCGGAGTCGGGAGCGGAAAGTCCATCTCAGCTCACACAGCAG	5461
QY	1821	VALGLIUNHISALAPROIETTPARGPROGLYTHRGUNINSERSESGLYSESGLYSER	1840
DB	5462	GTGAGACGACCACTTGTGAGACTGTGTACAGACGACACACGCGGAGCAGCGCGCAGC	5521

QY 1841 SerGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHis 1860
DB 5522 AGCGCGCGGGGTGGGGGACAGACGCGCGCGCGCTCCACCTCCACGACAC 5581
QY 1861 SerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsn 1880
DB 5582 TCGGCCATCTCCCTCGGAGCCAGAGATGCTCTCCAGCAGAGACCAGTGTCTTACACAC 5641
QY 1881 ThrGlyMetIysGlyIleIleThrAlaValGluProSerLysProThrValLeuArgSer 1900
DB 5642 ACAGAGCATGAAGGGTATCATACCGCTGTGAGCCAGACAGCCACGCTCTTAGGTCC 5701
QY 1901 ThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCysPro 1920
DB 5702 ACCCTCCACTCTCTCACTCCGCTGCGCCAGCTGCACATCTCCACTGCGCACCTGCCCCA 5761
QY 1921 LeuGlyGlyIleThrLeuAspGlyValTyrProThrIleuMetGluProValLeuLeuProLys 1940
DB 5762 CTGGGCGGACCCCTCGATGGGGGTCTACCTCACTGAGCGCGCTTGTGCTGCCAAG 5821
QY 1941 GluValArgProValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLeu 1960
DB 5822 GAGGCGCGCGGGGTGGCGCGCGCGCGAGCGCGCGCGAGCAGACCGCGCATGCTTCTC 5861
QY 1961 AlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGlu 1980
DB 5882 GCCAAGCGCCGAGCGCGCTCGGGCTGAGCGCGCTCTCTCCCGACAGAGGCTCGGAG 5941
QY 1981 ProArgProLeuValProProValSerGlyHisAlaThrIleAlaGlyThrProAlaLys 2000
DB 5942 CCGCGCGCGCTGAGCTCTCTGCTGCGCACCGCACCATCGCGCGCACCGCTCGCAG 6001
QY 2001 AsnLeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerLysAsp 2020
DB 6002 AACCTGCGACCTACACCGACCGCGCGCGAGCCCGCGCGCGCACTGCGCTCGGAC 6061
QY 2021 ProHisArgGlyIleGlySerSerLysArgProPheSerIleGlnIleuGluLeuArgSer 2040
DB 6062 CCGCACCGGGGAGAAAGTCAAGTAAAGCTTTTCCATCCAGAGAACTGGAATCTCGTTCT 6121
QY 2041 LeuGlyIleHisGlySerSerLysArgProGluGlyValGluProValSerProValSer 2060
DB 6122 CTGGGTTTACCGAGCGACAGCTACAGCCCGAAGGGGTGAGCCCGCTGACCGCTGTAGC 6181
QY 2061 SerProSerLeuThrHisAspLysGlyLeuProLysHisIleuGlnIleuAspLysSer 2080
DB 6182 TCACCCAGTGTGACCGACGACGAGGGGCTCCCAAGACCTGGAAAGGCTCGACAGAGC 6241
QY 2081 HisLeuGlnIleGluLeuArgProLysGlnProGlyProValIleuGlnIleGluAla 2100
DB 6242 CACTGGAGGGGAGCTGGCGCCCAAGCAGCAGCGCCCGTGAAGCTTGGCGGGAGGCC 6301
QY 2101 AlaHisIleuProHisIleuArgProLeuProGluSerGlnProSerSerSerProLeuLeu 2120
DB 6302 GCCCAGCTCCCAACCTCGCGCGCTGCTGAGAGCGCAGCTCTGTCACCGCTGCTC 6361
QY 2121 GlnThrAlaProGlyValIleGlyHisGlnArgValValThrLeuAlaGlnHisIleSer 2140
DB 6362 CAAACCGCGCGAGGGGTCAAGAGTCAACAGCGGGGTGTCACTTGGCGCGACATAGT 6421
QY 2141 GluValIleThrGlnArgIleHisIleProGlnIleuSerAlaProLeuPro 2160
DB 6422 GAGGTATCATCAGAGCTACCGGSCAACCCAGCAGAGAGTCAAGCGCACCGCTGCCC 6481
QY 2161 AlaProLeuIleSerPheProGlyAlaSerCysProValLeuAspLeuArgProPro 2180
DB 6482 GCCCGCTTACCTCTTCTCGGGGCGAGCTGCGCTCTTGAACCTCCCGCGCACCC 6541
QY 2181 SerAspLeuIleuProProProAspHisGlyAlaProAlaArgGlySerProHisSer 2200
DB 6542 AGTGACCTTCACTCCGCGCGCGGACCATGTGCCCCGCGCTGCTCCCCACAGC 6601
QY 2201 GluGlyGlyIleLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGluAsp 2220

DB 6602 GAAAGGGGCAAGAGTCTCCAGAGCCAAACAGAGCTCGGCTTGGGTGGTGAAGAC 6661
QY 2221 GlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAla 2240
DB 6662 GGTATTTAACTGTGTCCCAACCGGAGGGCATGAGCGAGCCAGGCACTCCCGAGTGTCT 6721
QY 2241 ValTyrProLeuLeuIleTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLys 2260
DB 6722 GTTATCCCGCTGCTGTACCGGATGGGGAACAGCGAGCCAGATGGCTTCCAG 6781
QY 2261 SerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSer 2280
DB 6782 TCTCCAGGCAACACCAACCGCGCGCGCGCTTCTTCAAGCAAGTGAACGAGAACACTCC 6841
QY 2281 AlaMetValIysSerLysGlnGlnIleAsnLysIleuAsnThrHisAsnArgAsn 2300
DB 6842 GCCATGGTCAAGTCAAGAAAGCAAGATCAACAAAGAGCTGAACACCCACACCGGAAAT 6901
QY 2301 GluProGluIleThrAsnIleSerGlnProGlyIleThrGluIlePheAsnMetProAlaIleThr 2320
DB 6902 GAGCTGAATACAAATATACAGCCAGCTGGAGCGAGATCTTCAATATGCCCCCATAC 6961
QY 2321 GlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnIleHisAlaSerThrAsnMet 2340
DB 6962 GGAACAGGCTTATGACTATGAGCCAGCGCGGTGAGAGAACATGCCACACCAACATG 7021
QY 2341 GlyLeuGlnAlaIleIleArgIleAlaLeuMetGlyLysTyrThrAspGlnTTPGluGluSer 2360
DB 7022 GGGCTGAGGCGCAATTAATTAAGAAAGCACTATGGTAAATATGACCTGGGAAAGTCC 7081
QY 2361 ProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerLysLeuProAlaAla 2380
DB 7082 CCGCGCTGACCGCCCAATGCTTTTAACTCTTAATCCATGCTGACGCTGCGCGCTGCT 7141
QY 2381 MetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGly 2400
DB 7142 ATGCCATATACCGCTGCTGAGCAGGAGTGAACACACTCACTGCGCAGGTGCGGC 7201
QY 2401 GlyIleAlaLysValSerGlyArgProSerSerArgIleValAlaLysSerProAlaProGly 2420
DB 7202 GGGAAAGGCCAAAGTCTTGTGGCAGACCCAGCAGCGGAAAGCCAAAGTCCCGCGCGC 7261
QY 2421 LeuAlaSerGlyAspArgProProSerValSerSerValHisSerGlnIleuAspCysAsn 2440
DB 7262 CTGGCATCTGGGACCGGCGCACTCTGTCTCTTCACTGATGACTCGAGGAGACTGCAAC 7321
QY 2441 ArgArgThrProLeuThrAsnArgValTTPGluAspArgProSerSerAlaGlySerThr 2460
DB 7322 CGCGGAGCGCGCTCAACCAACCGCGTGGGAGAGAGGCGCTGTCGCGAGGTTCCAGC 7381
QY 2461 ProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProPro 2480
DB 7382 CCAATCCCTTACAAACCCCGATCATGCGGTGACAGCGGGGTGTCAATGGCTTCCACCC 7441
QY 2481 ProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaTTPAspGlu 2500
DB 7442 CACCGCGGCTCCCGCGGCGAGCGGCGCTGCTGCGGCCCAACAGCTGGAGAG 7501
QY 2501 GluProLysProLeuLeuCysSerGlnTyrGluThrLysSerAspSerGlu 2517
DB 7502 GAGCCCAAGCACTGCTGCTGAGTACGAGACACTCCGACAGCGAG 7552

RESULT 2
AF125672 8686 bp mRNA 1linear PRI 04-APR-1999
LOCUS AF125672
DEFINITION Homo sapiens silencing mediator of retinoic acid and thyroid
hormone receptor extended isoform (SMRT) mRNA, complete cds.
ACCESSION AF125672
VERSION AF125672.1 GI:4559297
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

AAD 22973

REFERENCE 1 (bases 1 to 8686)
AUTHORS Park, E.-J., Schreier, D. J., Yang, M., Li, H., Li, L., and Chen, J. D.
TITLE SMRTe, a silencing mediator for retinoid and thyroid hormone receptors-extended isoform that is more related to the nuclear receptor corepressor
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (7), 3519-3524 (1999)
MEDLINE 9919215
PUBMED 10097068
REFERENCE 2 (bases 1 to 8686)
AUTHORS Chen, J. D.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-1999) Pharmacology and Molecular Toxicology, University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, MA 01655, USA
FEATURES
source Location/Qualifiers
1..8686
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..8686
/gene="SMRTe"
157..7680
/gene="SMRTe"
/function="transcriptional corepressor"
/note="SMRTe"
/codon_start=1
/product="silencing mediator of retinoic acid and thyroid hormone receptor extended isoform"
/protein_id="AAD2973.1"
/db_xref="GI:4559298"
/translation="MSGSTQVPAQVTRATEPRYPHSLSYVQIARTHTDGLLEVOH
HSDVASHLSQGSITGAPORRBSLSEPPGNERSGOELRLPESHVLYDELGSEMEF
ITSKPRRLILDPILRPSPLATGCPASSEDLTDRSLTGKLESPSPSPHTDDEL
ELVPRPLSKEBLQNDKRDREITWEOQISKLKKQOQLEBNAKPRPEKRVSP
ISSKRSIVQIYDENRKAKEAARILLEGQVVELPLYNQSDTRQYHENIKINDAM
RKKELIFGRBNHARKQMEQKFCORYDQLEAMEKKEVEIENPRRAKESKVEYE
KQPEIRKORELOERMQRVQGRSGLSMSPAREHEVESEIDELSGEENLEKQMDLA
VLPMLYKADDOORIKFINNGMLADPMKYKOROVNMMSSBOKEFRERFQHPQNF
GLIASLEKTAECVLYYLTQKNENYSILVRSYRRRSGKQOQOQOQOQOQOQOQOQ
QMPRSOEBKE
GRGTANSQGRKRIITRSMANEANSBEATTPQOASLAMELNESSRMTBEEVEITAKK
GLEHGRNWSAIAIARMGSKTVSQCKNFYVYKQMDLDELQHKLMKEKERNARKK
KAPAPASEBAAPVVEDEEMASGVTGNEEMVEAEATVANSDDTSISPSHTEA
ADTQNGKRPATLTGADGPPGPPPPPPEDIPAPTESPASPATLAPPPPPPPPP
SPRPVYPRKEKEKETAAAPRVEEREROKPAAEELAVDCKAEPRKSCTEBAEERG
AKGDAEABATLAKALKAKEKGGSGRTTAKSGAPODSDSATCSADVEDBAEQ
DKNRLLSPRPSLITPTGDPANASPOKPLDLKQIKORAAIPIQVTKVHBERDA
PKPAPAPPPPONLOPESDAPQPGSSPRGKSRAAPADKAERKPFPPAPAAEQ
KLPGDPCCWTSGLPFPVPPREVIKASPHADPSAFYAPGPHPLPLGLDTPAVALP
PPTISMPPLISAKHPSVLEQIGIAISQMSVTLQVPSSEHAKAVGPTWMLPLPM
DPRKLAPEGVQOEOQLSPRQAGPPRESLQVTPQOASVLRGTLGSGVSGSLTKGIS
TRVPSDAITTRGSLTHGTADVLYKGTIRITIGEDSPSLDGRSDSLPKHVIYEG
KKGHVLSTYEGMSVTQCKEDGSSGPPHETAPKRYDMEGRVGRASISASIGL
MGRAIIPERHSPHNLKOHINRISITQGIPIRSYVEAOBYLRKALILKREGEPP
PSDLTEAKTQALGLKLPKAEGLVATYKEAGRSIHETPREELHETPELPLAPPL
KGGSIQGTPIAKYDGTGASTTSGKKHVRSLISGRPTPPVHPLDVAADALAEACV
EESLKRRTGASSSGSIARGAIVYIPELCKPROSLIYEDHNAAPRAGLPPSPYTM
RETPRLQGSLSKASQDKRLTSTPRILAKSPHSTVPEHNHPISYVHILRGVSG
VDLRSHIPLAFDPTSIPIGRIPLDAAAAYLPHHILANPNYPLPYLIRLGRPTDA
LENRQTIINDYITSQOMHNATATAMAKQMDMLGSPRESLSLANTAAAGRGITIDISQ
VPLPLVLPPTPGTPATAMDRLAYLPTADPPFSRHSPLSLGCTHLTKPTTSS
BRERDRDRDRERERKSLITTTVENHPIWRPGEOSGSGSGSGSGSGSGSSPA
SHHAAQHSPISPRTDOLAQRSPSVLANTCMKGTITAVEKSKTVLRSTSSPVPA
ATPAPATHCPGLGTLGDTVFTLMEVYLPEKARVAPREPAPDTHGATLAKPPASG
LEPASSPSKSGSEPLVPVSGHATLARTAKNLAPHASPREPAPASLSDSHREKT
QSKPFIIOELIELSLGHSYSGBEQEPVSPVSPSLTIDDKLPHLELSDSHREKT
ELARKQPGVKGGEBAHLPHRLPESQSSSPLLOTAGVGVGHORVVTIAQHSIEV
IQQDTRHHPQOIASLPAVLSFPGASCVLLARPPSLIYLRPDHGAARGSHS
EGKRSRPERKTSVILGGEDGIEPVPBPBEMTEPGHSRSVYLYLRDQGTBSMG
SKSPNTSOPAPFISKLTBSNSAMVSKOELINKKMLDTHKRNPEPNTISQPGIEITNM
PATGTGLMTYRSQAVOEHAISTNMGLEAIIRKALMGKTYDQWESSPLSANAFPLNAS

ASLPAMPTITADGSDHTLTPGCGGAKAKVSGRRSSKAKSPAPGLASGDRPSSVS
VHSEDCNRRTPLTNRWEDPSSGSTPFPNPLMKLQAGVWASPPPPCLPAGSGP
LAGPHHAWDEEKPILCSQYENTLSDSE"

ORIGIN

Alignment Scores:
Pred. No.: 7,81e-254 Length: 8686
Score: 12978.00 Matches: 2487
Percent Similarity: 98.57% Conservative: 3
Best Local Similarity: 98.46% Mismatches: 9
Query Match: 98.21% Indels: 28
DB: Gaps: 3

US-09-522-753-5 (1-2517) x AFL25672 (1-8686)

QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro 20
157 ATGTGGGCTCCACACAGCCTTGCGACAGAGCTGGAGGGCCACGTGACCCGCTACCGG 216
Db 21 ProHisSerLeuSerTyrProValGlnIleAlaGlnThrHisThrAspValGlyLeuLeu 40
217 CCCACAGCCTTCTTCCACCCAGTGAAGATCGCCGACCCGACACGACGACCTCGGCTCTG 276
QY 41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisIleLeuSerProGlySerIleIleGln 60
277 GAGTACCAAGACACACTCCCGACATGACTCCACCTTGGCCGCTCCATCATCCAG 336
Db 61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyValAsnGluArgSerGln 80
337 CCCACGCGCGGAGGCGCTCCCTGCTGCTGAGTTCCACCCGGGAATGAAGCGTCCAG 396
QY 81 GluLeuHisIleLeuArgProGluSerHisSerTyrLeuProGluLeuGlyIleSerGluMet 100
397 GAGCTCCACCTCGCGGCAAGTCCCATCATCTGCTCCGACCTGGGAAGTCAAGAGATG 456
Db 101 GluPheIleIleGluSerTyrAspProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
457 GAGTTCATTGAAAGCAAGCGCCTCGCTAGAGCTGCTCTGACCCCTCTGCTGCAACG 516
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrIleAspArgSer 140
517 TACACCTCTGCTGCGACGCGGCGACGCTGCGGATCTGAAAGACTTCAAGAGACCGTAGC 576
QY 141 LeuThrGlyIleLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
577 CTGACCGGCGAAGCTGAAACGGGTGCTCCCCACGCCCCCGACACTGACCTGAGCTG 636
Db 161 GluLeuValProProArgLeuSerTyrGluGluLeuIleGlnAsnMetAspArgValAsp 180
637 GAGCTGCTGCGCCACGCTGCTCAAGAGAGAGCTGATCCAGACATGAGCCCGTGGAC 696
QY 181 ArgGluIleThrMetValGluGlnGlnIleIleSerTyrLeuIleIleGlnGlnGlnLeu 200
697 CGAGAGATCAACATGTTAGAGACAGACGATCTGAAGCTGAAGAGAGAGAGAGAGCTG 756
Db 201 GluGluGlnAlaAlaIleAspProProGluProGluIleAspValSerProProIleGlu 220
757 GAGAGAGAGAGCTGCCAAGCGCGCCGAGCTGAGAGACCCGCTGACCGCGCCATCGAG 816
QY 221 SerTyrHisArgSerLeuValGlnIleIleTyrAspGluAsnArgIleValIleAla 240
817 TCGAAGACACCGGACCTGGGTCGATCACTACGAGAGAACCGGAGAGAGCTGAAGCT 876
Db 241 AlaHisArgIleLeuGluGlnIleGluGlyProGlnValGluLeuProLeuTyrAsnGlnPro 260
877 GCACATCGGATCTGGAAGCGCTGGGCGCCGAGTGAAGCTGCGCTGTAACACGAGCCC 936
QY 261 SerAspThrArgGlnTyrHisGluAsnIleIleValIleAsnGlnAlaMetArgIleValLeu 280
937 TCGGACACCGGACGATCATGAGAACATCAAAATTAACACGAGCGATGCGGAAGAGCTA 996
Db 281 IleLeuTyrPheIleArgArgArgHisAlaIleArgIleGlnTyrIleGlnIleValPheGlyGln 300

Db 997 ATCTTGATCTTCAAGAGAGGAATCAAGCTCGAAGAAACAATGGAGACAGAACTTCTGCAG 1056
Qy ArgTyraSpGlnLeuMetGluAlaLeuGluTyrsValGluArgTlleGluAsnAspPro 320
Db 1057 CGCTATACCAAGCTCATGGAGGCTTGGAGAAAGAGTGGAGCCGATCGAAGAACACCC 1116
Qy ArgTargTargAlaTyrsGluSerTyrsValArgTluTyTyrGluTyrsGlnPheProGluTle 340
Db 1117 CGGCGCGGCGCAAGAGAGCAAGTTCCGAGTACTACGAAAGACAGTTCCCTGAGATC 1176
Qy ArgTyrsGlnArgGluLeuGlnGlnLeuArgMetGlnSerArgValGluGlnArgTlleSerGly 360
Db 1177 CGCAAGAGCGCGGCTGCAGAGCGCATGCGAG---AGGGTGGCGCACGCGGCGCATGCGG 1233
Qy LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluTlleLeuAspGlyLeuSer 380
Db 1234 CTGTCCATGTGCCCCCGCCGAGCGAGACAGAGTGTTCAGAGTATCATGAGGCTCTCA 1293
Qy GluGlnGlnLeuAsnLeuGluTyrsGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400
Db 1294 GAGCAGAGAACTCGAAGACAGATGCGCCAGCTGGCCGTGGATCCCGCCCATCTGTAC 1353
Qy AspAlaAspGlnGlnArgTlleTyrsPheTleAsnMetAsnGlyLeuMetAlaAspProMet 420
Db 1354 GACGCTACCAAGCGCCCATCMAGTTCAACAATGAACGGGCTTATGGCCGACCCCATG 1413
Qy TyrsValTyrTyrsAspArgGlnValMetAsnMetTyrSerGluGlnGluTyrsGluThrPhe 440
Db 1414 AAGGTGTACAAAGACCCCGAGGTATGAACATGTGTGATGAGCAGAGAGAGAACCTTC 1473
Qy ArgGluTyrsPheMetGlnHisProTyrsAsnPheGlyLeuTlleAlaSerPheLeuGluArg 460
Db 1474 CGGAGAAAGTTACATGACAGCATCCCAAGAACTTTGGCTGATCCGATCATCTCTGAGAGG 1533
Qy TyrsThrValAlaGluCysValLeuTyrTyrTyrLeuThrTyrsTyrsAsnGluAsnTyrTyrs 480
Db 1534 AACAAGTGGCTGAGTGCCTCTTATTACTGACTGAAGAAATGAGAACTATAG 1593
Qy SerLeuValArgArgSerTyrArgArgGluTyrsSerGlnGlnGlnGlnGlnGlnGln 500
Db 1594 AGCTGTGTGAGAGAGCTATCGCGCGCGCAAGAGCCAGCAGCAACACGACAG 1653
Qy GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
Db 1654 CACGAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1713
Qy AspGluTyrsGluTyrsGluTyrsGluAlaGluTyrsGluGluTyrsProGluValGluAsn 540
Db 1714 GATGAGAAAGAGAAAGAAAGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1773
Qy AspTyrsGluAspLeuLeuTyrsGluTyrsThrAspAspThrSerGluValAspAsnAspGlu 560
Db 1774 GACAAAGAAACCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1833
Qy TyrsGluAlaValAlaSerTyrsGluArgTyrsThrAlaAsnSerGlnGlnGlnGlnGlnGln 580
Db 1834 AAGAGGCTGTGGCTTCCAAAGCCGCAAAATCTCCAAACAGCCAGGAAACCCAAAGC 1893
Qy ArgTlleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaTleThrProGlnGln 600
Db 1894 CGCATCAACCCGCTCAATGGCTAAATGAGGCCAAACAGCAGAGAGGCAATCAACCCCAAGCAG 1953
Qy SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTyrThrThrGluGluGluMet 620
Db 1954 AGGCGGAGGCTGGCTTCATGAGAGCTGAATGAGAGTTCTGCTGAGAGAGAGAGAGAGATG 2013
Qy GluThrAlaTyrsTyrsGluLeuLeuGluHisGluArgTyrAsnThrPheSerAlaIleAlaArgMet 640
Db 2014 GAAAGACGCAAGAAAGGTCTCTGGAACAAGCGCGCAACTGTGCGCACTGCGCCGATG 2073
Qy ValGlySerTyrsThrValSerGlnCysTyrsAsnPheTyrPheAsnTyrTyrsTyrsGln 660
Db 2074 GTGGGCTCCAAAGCTGTGTGAGTGAAGAACTTCTCACTTCAAGAGAGAGAGCAG 2133

Qy AsnLeuAspGluTlleLeuGlnGlnHisTyrsLeuTyrsMetGluTyrsGluArgAsnAlaArg 680
Db 2134 AACTCATGATGATCTTTCAGACAGCACAAGCTGAAGATGAGAAAGAGAGAGAGAGAGAG 2193
Qy ArgTyrsTyrsTyrsValProAlaAlaSerGluGluAlaAlaPheProProValVal 700
Db 2194 AGGAAAGAAAGAAAGAGCGCGCGCGCGCGAGAGAGAGAGGTGATTCGCCGCGTGGTG 2253
Qy GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluGluMetValGluGlu 720
Db 2254 GAGGATAGAGAGATGAGAGCGGTGCGGCGGTGACGGGAAATGAGAGAGATGTGAGAGAG 2313
Qy AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGlyCysSerGlyProAla 740
Db 2314 GCTGA-----GCC 2322
Qy ThrValAsnAsnSerSerAspThrGluSerTleProSerProHisThrGluAlaAlaTyrs 760
Db 2323 ACTGTCAACAAAGCTCAGACACCGAGAGCATCCCTCTCTCACACTGAGGCGCCAG 2382
Qy AspThrGlyGlnAsnGlyProTyrsProProAlaThrLeuGluValAspGlyProProPro 780
Db 2383 GACACAGGCAAGATGGGCCCAAGCCCAAGCCACTGGGCGCGGAGCCCA 2442
Qy GlyProProThrProPro-ArgArgThrSerArgAlaProIleGluProThrProAlaSer 800
Db 2443 GGGCCACCCACCCCAACCGAGAGACATCCCGCC-CCCATGATGTCCACCCGCGCTC 2501
Qy ArgAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProProPr 820
Db 2502 TGAAGCACCTTATGGCCCTTATGGCCCGCCACAGCACCCCATTTCTCTTACCTCTCC 2561
Qy OValValProTyrsGluGluTyrsGluGluGluGluGluGluGluGluGluGluGluGlu 840
Db 2562 TGTGTCTCCCAAG 2621
Qy TyrsGluGluGlnTyrsProProAlaAlaGluGluLeuAlaValAspThrGlyTyrsAlaGlu 860
Db 2622 GAGAGAGCAAG 2681
Qy UProValTyrsSerGluCysThrGluGluGluGluGluGluGluGluGluGluGluGlu 880
Db 2682 GCCGTCAAG 2741
Qy ArgAlaAlaGluAlaThrAlaGluGluAlaLeuTyrsAlaGluTyrsGluGluGluGlu 900
Db 2742 GGAAGCGCTGAG 2801
Qy ArgTyrsArgAlaThrThrAlaTyrsSerSerGlyAlaProGlnAspSerAspSerAlaThr 920
Db 2802 CGGCGAGGCGCACCAAG 2861
Qy TyrsSerAlaAspGluValAspGluAlaGluGluGluGluGluGluGluGluGluGluGlu 940
Db 2862 CTGCAAGTGAAG 2921
Qy OArgProSerLeuLeuThrProThrGluAspProAlaAlaAsnAlaSerProGluTyrsPr 960
Db 2922 AAGGCCAGGCTCTCAACCCGACTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2981
Qy OLeuAspLeuTyrsGluLeuTyrsGluGlnArgAlaAlaAlaIleProProIleGlnValThrTy 980
Db 2982 ACTGACCTGAACAGTGAAG 3041
Qy sValHisGluProProArgGluAspAlaAlaProThrTyrsProAlaProProAlaProPr 1000
Db 3042 AGTCCATGAGCCCGCGGAG 3101
Qy OProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProAr 1020
Db 3102 GCCACCGCAAAACCTGACAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3161

1020 gGlylySserArgSerProAlaProProAlaAspLySglu----- 1033
Db 3169 GGGCAAGAGCAGAGGCCGGGACCCCGCCGACAAAGAGGAGAGAGCCGTGTCTT 3221
Qy 1034 ----AlaPheAlaAlaGluAlaGlnLysLeuProGlyVaAspProProCysTrpThsSerG 1052
Db 3222 CCAGAGCTTCGAGCGAGCGAGCCAGAAAGTCCTGGGAGACCCCTTGTCGACCTTCGG 3281
Qy 1052 yLeuProPheProValProProArgGluValIleLysAlaSerProHisAlaProAspPr 1072
Db 3282 CTTGCGCTTCCCGTCGCGCCCGCTGAGGTGATCAAGGCTCCCGCATGCGCCGAGACC 3341
Qy 1072 oSerAlaPheSerTyAlaProProGlyHisProLeuProLeuGlyLysHisAspThrAl 1092
Db 3342 CTCAGCTTCTCTTAAGCTCACTGTCACCTGATCAACCCAGCTCCCTGGCTCATGACACTGC 3401
Qy 1092 aArgProValLeuProArgProProThrIleSerAsnProProProLeuIleSerSerAl 1112
Db 3402 CCGGCGCGTTCCTGCGCGCCACCCACCATCTCCAAACCGGCTCCCTCATCTCCCTGC 3461
Qy 1112 aLysHisProSerValLeuGluValIleGlyAlaIleSerGlnGlyMeSerValG 1132
Db 3462 CAAGCACCCAGCGCTCTCGAGAGGCAATAGGTGCATCTCCAAAGAAATGCGGTCA 3521
Qy 1132 nLeuHisValProTySerGluHisAlaLysAlaProValGlyProValThrMetGlyLe 1152
Db 3522 GCTCCAGCTCCGTACTCAGAGCATGCTCAAGGCCCGGTGGGCTGTCAACCATGGGGCT 3581
Qy 1152 uProLeuProMetAspProLysAlaLeuAlaProPheSerGlyValLysGlnGlnIle 1172
Db 3582 GCCCTGCGCATGAGCCCAAAAGCTGCACCTTCAGCGAGGTAGACGAGAGAGAGCT 3641
Qy 1172 uSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGlnAl 1192
Db 3642 GTCCCAACGGGGCCAGGCTGGGCCACCGAGAGCTGGGGGTGCCACAGACCCAGAGGC 3701
Qy 1192 aSerValLeuArgGlyYThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyI 1212
Db 3702 GTCCGCTGAGAGGAGACAGCTGTGGGCTCAGTTCCGGGGAGAAAGCATCAACAAAGCAT 3761
Qy 1212 eProSerThrArgValProSerAspSerAlaIleThrTyArgGlySerIleThrIleG 1232
Db 3762 TCCACAGACACGGGTCCCTCGGACAGCGCATCATACCGGGCTCCATCCACACG 3821
Qy 1232 yThrProAlaAspValLeuTyLysGlyYThrIleThrArgIleIleGlyGluAspSerPr 1252
Db 3822 CACGCCAGCTGAGCTCTGTACAGAGGACCATCATCAGGATCATGGCGAGAGACAGCC 3881
Qy 1252 oSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyGlnG 1272
Db 3882 GAGTCGCTTGAGACCGGGCGGGAGAGACAGCTGCCCAAGGGCACGTCATCTACGAAAG 3941
Qy 1272 yLysLysGlyHisValLeuSerTyGlnGlyMeSerValThrGlnCysSerTySerG 1292
Db 3942 CAAAGAGGGGCAAGCTTGTCTATGAGGGTGCAATGTGTGTAACCAAGTCCTCAAGGA 4001
Qy 1292 uAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyArg 1312
Db 4002 GAGCGGCAAGAGAGCTCAGAGACCCCGCATGAGAGCGCGCCCGCAAGGACCTATGA 4061
Qy 1312 PheMetGlnGlyArgValGlyValArgAlaIleSerSerAlaSerIleGlnGlyLeuMetG 1332
Db 4062 CATGATGAGAGGCGCGGTGGGAGAGCATCTCTCAGCGACGATCGAAAGTCTCATGGG 4121
Qy 1332 yArgAlaIleProProGluArgHisSerProHisIleLysLeuLysGlnGlnHisIleAla 1352
Db 4122 CCGTGCATCTCCCGCGAGAGACACAGCCCGCACCTCAAGAGAGAGACACACATCCG 4181
Qy 1352 gGlySerIleThrGlnGlyLeuProArgSerTyValGlnAlaGlnGluAspTyLeuArg 1372
Db 4182 CGGGTCATCATCAAGAGAGATCCCTGGTCTCATGTAAGGACACAGAGAGATACCTGGC 4241
Qy 1372 gATGAlaAlaLysLeuLysArgGlnGlyYThrProProProProProSerArgAs 1392

4242 TCGGAGGCGCAAGCTCTTAACCGGAGAGGACGGCTCCGCCCCACGCGCTCATCGGGA 4301
Qy 1392 PleuThrGlnAlaTyLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisG 1412
Db 4302 CTTGACCGAGGCTTACAGAACCGCAGGCCCTGGGCCCTCGAAAGCTGAAAGCCGCCATGA 4361
Qy 1412 uGlyLeuValAlaThrValLysGlnAlaGlyValSerSerIleHisGlnIleProArgGln 1432
Db 4362 GGGCTGTGGTGCACAGGTGAAGAGGCGGCGCTCATCATGAGATCCCGCGCAGGA 4421
Qy 1432 uLeuArgHisThrProGluLeuLeuProLeuAlaProArgProLeuLysGlnGlySerIleTh 1452
Db 4422 GCTGGGACACAGCCCGAGCTGCTCCCTGGCGCGCTCAAGAGAGGCTCATAC 4481
Qy 1452 rGlnGlyThrProLeuLysTyArgPThrGlyAlaSerThrThrGlySerLysLysHisAs 1472
Db 4482 GAGGGGACCCCGCTCAAGTACAGACCGGCCGCTCACCACTGGCTCCAAAAGACAGA 4541
Qy 1472 pValArgSerLeuIleGlySerProGlyValArgThrPheProProValHisProLeuAspVa 1492
Db 4542 CGTAGCTCTCTCATTCGGACGCCCGCGGAGAGTCCACCCGTCACCCGCTGAGATGT 4601
Qy 1492 lMetAlaAspAlaArgAlaLeuGluValaCysTyArgGlnGluSerLeuLysSerArgPr 1512
Db 4602 GATGGCCGAGCGCCGGGCACTGAAAGTGCCTGCTACGAGAGAGCTTAAGAGCCGAGCC 4661
Qy 1512 oGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProG 1532
Db 4662 AGGAGCCGCGACAGCTCGGGGGCTCCATTGCGCGGCGCCGCGTCAATTGTGCTCA 4721
Qy 1532 uLeuGlyLysProArgGlnSerProLeuThrTyGlnAspHisGlyAlaProPheAlaG 1552
Db 4722 GCTGGGTAGCCGCGGACAGAGCCCTGTGACTATGAGACACAGGGGACCCCTTGCCGG 4781
Qy 1552 yHisLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGln 1572
Db 4782 CCACTCCACAGAGGTTCGCGGTGACCATGGCGGAGCCAGCCGCGCTGAGAGAG 4841
Qy 1572 ySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGlnI 1592
Db 4842 CAGCTTTCCTCAGAGAGCATCCAGAGACGAAAGCTGACGTGACCGCTGTGAGAT 4901
Qy 1592 eAlaLysSerProHisSerThrValProGluHisIleProHisProIleSerProTyG 1612
Db 4902 CGCAAGTCCCGGACAGACCGTGGCCGAGACACCCACATCCCATCTCGCCATGA 4961
Qy 1612 uHisLeuLeuArgGlyValSerGlyValaAspLeuTyArgSerHisIleProLeuAlaPh 1632
Db 4962 GCACTGCTTGGGGGTGAGTGGCGGTGACCTGTACCGACGACATCCCTGGGCTT 5021
Qy 1632 eAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyArgTyLeuPr 1652
Db 5022 CGAGCCCATCTCATACCCCGGGGATCCCTTGAGAGAGCGCGCTCATCACTGGCC 5081
Qy 1652 oArgHisLeuAlaProAsnProThrTyProHisIleuTyProProTyLeuIleArgG 1672
Db 5082 CGAGACACTGGCGCCCAACCCACCTACCGGACCTGTACCACTTACCTCGCGCG 5141
Qy 1672 yTyProAspThrAlaAlaLeuGluAspArgGlnIleIleAsnAspTyIleThrSe 1692
Db 5142 CTACCCGACAGCGCGCGCTGGAGAACCGGACAGACATCATCAATGATCATCACTC 5201
Qy 1692 rGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArg 1712
Db 5202 GAGAGATGACACACACAGCGCCACCGCATGGCCGACGAGCATGATGTGAGGGG 5261
Qy 1712 yLeuSerProArgGluSerSerLeuAlaLeuAsnTyAlaAlaGlyProArgGlyIleI 1732
Db 5262 CCTCTGCGCGGAGATCTCGCTGGCACTCACTACGCTGCGGGTCCCGAGAGCATCAT 5321
Qy 1732 eAspLeuSerGlnValProHisIleuProValLeuValProProThrProGlyYThrProAl 1752

Db 5322 CGACCTGCCAAGTCCACACCTGCTGCTGCTGCCGCCGACACCAAGGACCCCAACG 5381
Qy 1752 aThrAlaMetAspArgLeuAlaTyrLeuProThraAlaProGlnProPheSerSerArgH 1772
Db 5382 CACGGCATGGAGCGGCTTGCTGCTACCTCCACCGCGGCCCAAGCCCTTCAGACACCGGCA 5441
Qy 1772 sSerSerSerProLeuSerProGlyGlyProThiHleuThriLysProThrThrHse 1792
Db 5442 CAGAGCTCCCTCCTCTCCCGAGGAGGTCCAACACACTTGCAAAACCAACACCAACGTC 5501
Qy 1792 rSerSerGluArgGluArgAspArgAspArgGluArgAspArgGluArgGluArg 1812
Db 5502 CTGCTCCGAGGGAGGAGCGGAGCCGGAGTCGAGCGGGAGCCGGAGTCGGAGCGGAGAA 5561
Qy 1812 sSerLeuLeuThSerThrThrThrValGluHleAlaProIleThraArgProGlyThrG 1832
Db 5562 GTCCATCTCACTGCTCCACCAAGCGGTGGAGCAGCAACCTCTGGAGACTGGTACGA 5621
Qy 1832 uGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgProAl 1852
Db 5622 GCAGAGCAGCGGCGAGCAGCGGCGGCGGGGTGGGGGCGAGCAGACCGCCCGC 5681
Qy 1852 aSerHisSerHisAlaHleGlnHisSerProIleSerProArgThrGlnAspAlaLeuG 1872
Db 5682 CTCCCATCTCCATGCTCCACCAAGCAGCACTGCCCATCTCCCTCGAGCCAGATGCTTCA 5741
Qy 1872 nGlnArgProSerValLeuHleAsnThrGlyMetLeuGlyIleIleThraAlaValGluPr 1892
Db 5742 GCAGAGACCCAGTGTCTTCAACAACAGGCAATAGAGGATTCACCGCTGTGGAGCC 5801
Qy 1892 oSerLysProThrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaTh 1912
Db 5802 CACCAAGCCCAAGGCTCTGAGTCCACTCACTCCCTCAACCGTTCGCCAGCTGCAC 5861
Qy 1912 rPheProProAlaThrHisGlyProLeuGlyGlyThLeuAspGlyValIleProThrLe 1932
Db 5862 ATTCCACCTGCAACCACTGCGCACAGCGGCGGACCTCGATGGGGTCTAACCTTACCT 5921
Qy 1932 uMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProAr 1952
Db 5922 CATGGAGCCGCTTGTGCTGCCAAGAGGCCGCCCGGGTGGCCGCGCAAGCGGCCCCG 5981
Qy 1952 gAlaAspThrGlyHleAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAl 1972
Db 5982 AGCAGACACCGGCGCATGCTTCTCGCCAGCCGCCCGCGCTCGGGCTGGAGCGCCG 6041
Qy 1972 aSerSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHleAl 1992
Db 6042 CTCTCTCCCGCAGAGGGCTCGAGGCCCGGCCCTTAGTGTCTGTCTGTGGCCACGC 6101
Qy 1992 aThrIleAlaArgThrProAlaLysAsnLeuAlaProHleHisAlaSerProAspProP 2012
Db 6102 CACCACTGCCCGGACCCCTGCGAAGAACTTCGCACTTCAACAGCCGCGGACCGGCC 6161
Qy 2012 oAlaProProAlaSerAlaSerAspProHleArgGluLysThrGlnSerLysProPhe 2032
Db 6162 GGGGCACTGCTCGGCTCGGACCGGACCGGAAAGACCTCAAGTAAACCTTTTC 6221
Qy 2032 rIleGlnGluLeuGluLeuArgSerLeuGlyIleHleGlySerSerLysProGluGlu 2052
Db 6222 CATTCCGAACTGGAACTCGGTCTCTGGGTTCACACGCGCAGCAGTACAGCCCGAAG 6281
Qy 2052 yValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLys 2072
Db 6282 GGTGGAGCCCGTACGCTGTGAGCTCACCACTTGACCCACAGCAAGAGGCTCCCCA 6341
Qy 2072 sHisLeuGluGluLeuAspLysSerHisLeuGluGluGluLeuArgProLysGlnProG 2092
Db 6342 GCACCTGGAAGACTCGACAAAGACCACTGAGGGGAGCTGGGCGCAAGAGCAGG 6401
Qy 2092 yProValLysLeuGlyGlyGluAlaAlaHisLeuProHleAspArgProLeuProGlySe 2112
Db 6402 CCCGTTAAGCTTGGCGGGAGGCGGCCACCTCCACACCTGCGGCTGTGCTGAGAG 6461

Qy 2112 rGlnProSerSerSerProLeuLeuGlnThraAlaProGlyValLysGlyHisGlnArgVa 2132
Db 6462 CCAGCCCTCGTCCAGCCCGGCTGCTCCAGACCGGCCCGAGGGGTCAAAAGTTCACAGCGGCT 6521
Qy 2132 lValThLeuAlaGlnHisIleSerGlyValIleThGlnAspTyrThrArgHisIlePr 2152
Db 6522 GGTCACTTGCGCCAGCAGCATCAGTGAAGTATCACAAGAGTATCAACCCGGCACACCC 6581
Qy 2152 oGlnGlnLeuSerLysProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysPr 2172
Db 6582 ACAGCACTCAAGCGCACCCCTGCGCCCGCTTACTCTCTTCTGGGGCCACTGCCCC 6641
Qy 2172 oValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAlaAspHisGlyAl 2192
Db 6642 GTCTCTGACCTCCGCGGCCCAACCATGACCTTACTCCGCGCCCGGACCATGGTGC 6701
Qy 2192 aProAlaArgGlySerProHleSerGlyGlyGlyLysArgSerProGluProAsnLysTh 2212
Db 6702 CCGGCGCGTGGCTCCCCCAAGCGAAGGGGGCAAGAGGTCTCCAGAGCCCAACAGAC 6761
Qy 2212 rSerValLeuGlyGlyGlyLysAspGlyIleGluProValSerProProGluGlyMetTh 2232
Db 6762 GTCGCTTGGAGTGGTGAAGACGGGTATTGAACCTGTGTCCCAACGAGGCGCATGAC 6821
Qy 2232 rGluProGlyHisSerArgSerAlaValIleProLeuLeuTyrArgAspGlyGluGlnTh 2252
Db 6822 GAGACCGGCACTCCCGGAGTGTGTGTAACCGCTCTGTATCCGGAGTGGGAAACAGAC 6881
Qy 2252 rGluProSerArgMetGlySerLysSerProGlyAsnThSerGlnProProAlaPhePh 2272
Db 6882 GAGAGCCAGAGATGGGTCTCCAAGTCTCCAGGCAACACGACGCGCAGCTTCTT 6941
Qy 2272 eSerLysLeuThrGluSerIleAsnSerAlaMetValLysSerLysLeuGlnGluIleAsnLy 2292
Db 6942 CACGACCTCAACGAGAGCAACTCCGCAATGCTCAAGTCCAAAGAAAGAAAGATCAACAA 7001
Qy 2292 sLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrG 2312
Db 7002 GAAGCTGAACCCCAACACCGGAAATGAGCTGAAATACAAATACAGCCAGCTGGAGCGA 7061
Qy 2312 uIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaVa 2332
Db 7062 GATCTTCAATATATCCGCGCATCAACGGAAACAGGCTTATGACCTATGAAAGCCAGCGGT 7121
Qy 2332 lGlnGluHleAlaSerThrAsnMetGlyLeuGluAlaIleIleArgGlyAlaLeuMetG 2352
Db 7122 GCAGAAACATGTCAGACCAACATGGGGCTGAGGCGCATTAATTAAGAAAGCACATGAGG 7181
Qy 2352 yLysTyrAspGlnTTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAs 2372
Db 7182 TAAATATGACAGTGGGAAGAGTCCCGCGCTCAAGCGCCAAAGCTTTTAACTCTGAA 7241
Qy 2372 nAlaSerAlaSerLeuProAlaAlaMetProIleThraAlaAlaAspGlyArgSerAspHis 2392
Db 7242 TGCCAGTGCAGCGCTCGCGCTGTATGCCATTAACGCTGAGAGGAGTGAACA 7301
Qy 2392 sThrLeuThrSerProGlyGlyGlyGlyLysValAluValSerGlyArgProSerSerAr 2412
Db 7302 CACACTCACTTCGACAGTGGCGGCGGAAAGGCAAGGTCTTCGGCAGACCCAGCAGCGG 7361
Qy 2412 gLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSe 2432
Db 7362 AAAAGCAAGTCCCGCGGCCCGGCGCTGGCATTTGGGAGCCGCGCACCTCTGTCTCTC 7421
Qy 2432 rValHisSerGluGlyLysProCysAsnArgArgThrProLeuThrAsnArgValTTrpGluAs 2452
Db 7422 AGTGACACTCGAGAGGAGACTGCAACCGCGGAGCGGCTCAACCAACGCGTGTGGAGGA 7481
Qy 2452 pArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuHisMetArgLeuG 2472
Db 7482 CAGGCGCTGTGCGAGGTTCCAGGCAATTCCTTCAACACCCCTGATATCGGCTGCA 7541

QY 2472 nAaGlyValMetAlaSerProProGluLeuProAlaGlySerGlyProLeuAla 2492
| | | | |
Db 7542 GCGGGGTGTCATGGCTTCCCAACCCCAACGGGGCCCTCCCGGGGAGCGGCCCTCGC 7601
QY 2492 AGlyProHisHisAlaATPASPGLUGLUProLysProLeuLeuCySerGlnTyrGln 2512
| | | | |
Db 7602 TGGCCCCCAACCGCTGGGAGAGAGAGCCAAACCACTGCTGTCTGCAAGTACAGAAC 7661
QY 2512 rLeuSerASPserGlu 2517
| | | | |
Db 7662 ACTCTCCGACAGCGAG 7677

RESULT 3
AR447713 9053 bp DNA linear PAT 20-FEB-2004
LOCUS Sequence 306 from patent US 6673549.
DEFINITION AR447713
ACCESSION AR447713 GI:42676037
VERSION AR447713.1 GI:42676037
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 9053)
AUTHORS Furness, L.M. and Buchbinder, J.L.
TITLE Genes expressed in C3A liver cell cultures treated with steroids
JOURNAL Patent: US 6673549-A 306 06-JAN-2004;
FEATURES Location/Qualifiers
source 1. 9053
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Alignment Scores:

Pred. No.: 1 09e-252 Length: 9053
Score: 12922.00 Matches: 2487
Percent Similarity: 98.26% Conservative: 2
Best Local Similarity: 98.18% Mismatches: 10
Query Match: 97.78% Indels: 35
DB: 6 Gaps: 5

US-09-522-753-5 (1-2517) x AR447713 (1-9053)

QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTPArgAlaThrGluProArgTyrPro 20
| | | | |
Db 520 ATGTGGGATCCACACAGCTGTGGCAGACGTGAGGGCCACTGAGCCCGCTACCCG 579
QY 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
| | | | |
Db 580 CCCACAGGCTTCTTCCATCCAGTGCAGATCGCCCGACACACGACGCTCGGGCTCTG 639
QY 41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
| | | | |
Db 640 GAGTACACACACTCCCGGACATATGCTCCCACTGTCCCGGCTCCATCATCCAG 699
QY 61 ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
| | | | |
Db 700 CCCACGGCGGAGGCGCTCCCTGCTGTGAGTTCCAGCCCGGAGATGAAGCTCCAG 759
QY 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGlyLeuGlyLysSerGluMet 100
| | | | |
Db 760 GAGCTCCACCTGGGCGCAAGTCCCACTCATCTCCGAGCTGGGAGATCAGAGATG 819
QY 101 GluPheIleGlySerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
| | | | |
Db 820 GAGTTTATTGAAAGCAAGGCGCTCGGCTHAGCTGTGCTGACCCCTGCTGCAACG 879
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
| | | | |
Db 880 TCAACCCCTGCTGGCCACGGGGCCAGCTGGGGATCGAAGACTCAACCAAGACCGTAG 939
QY 141 LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
| | | | |
Db 940 CTGACGGGCGAAGCTTGAAACCGGTGTCTCCCGCCAGCCCGCCGACACTGAGCTG 999

QY 161 GluLeuValProProArgLeuSerLysGluLeuIleGlnAsnMetAspArgValAsp 180
| | | | |
Db 1000 GAGCTGTGGCCGCCACAGGCTGTCCAAAGGAGAGCTGATCCAGAAACATGACCGCGTGAC 1059
QY 181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeu 200
| | | | |
Db 1060 CGAGAGATCAACCATGTAGAGAGAGAGATCTTAACTGTAAAGAAAGAGAGACAGACGTG 1119
QY 201 GluGluGluAlaAlaLysProProGluProGluLysProValSerProProIleGlu 220
| | | | |
Db 1120 GAGGAGAGGCTGCCAAGCCGCCGAGCTGAGAAAGCCGTGTACCGCGCCCATTCAG 1179
QY 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysValAlaGluAla 240
| | | | |
Db 1180 TCGAAGACCCGCAAGCTGTGGTGCAGATCTACGACGAGAAACCGAAGAGCTGAGCT 1239
QY 241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrTrpAsnGlnPro 260
| | | | |
Db 1240 GCACATCGGATTTCTGAAAGGCTGGGGCCCGAGGTGAGAGCTGCCCTGTACCAACCGCC 1299
QY 261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
| | | | |
Db 1300 TCCGACACCCGCACTATCATGAAACATCAAAATTAACCGCATGCGGAAGAGCTA 1359
QY 281 IleLeuTyrPheLysArgArgHisHisAlaArgLysGlnTyrPlyGlnLysPheCysGln 300
| | | | |
Db 1360 ATCTTGTACTTCAAGAGAGAGAAATCACGCTCGGAAACAATGGAGACAGAAATTTGCCCAG 1419
QY 301 ArgTyrAspGlnLeuMetGluAlaLeuGlyLysLysValGluArgIleGluAsnAsnPro 320
| | | | |
Db 1420 CGCTATGACCACTCATGAGAGGCTGGGAGAAAGTGGAGCGCATTCACAAACACCCC 1479
QY 321 ArgArgArgAlaLysGlySerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340
| | | | |
Db 1480 CCGCGGCGGGCCAAAGAGAGAGAGGTGCCGAGTCTACGAAAGACAGTTCCTTGAGATC 1539
QY 341 ArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGly 360
| | | | |
Db 1540 CGCAAGCAGCGGAGCTGCAGAGCGCATGACAGACAGGCTGGCGCAGCGGGCACTGTGG 1599
QY 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
| | | | |
Db 1600 CTGTCCATGTCCGCGCCCGCAGCAGACGAGGTGTCAAGATTCATGATGCTCTCA 1659
QY 381 GluGlnGluAsnLeuGlnLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400
| | | | |
Db 1660 GAGCAGAGAGACCGGAAGAGAGATGCCAGCTGGCCGTGATCCCGCCCATGCTGATC 1719
QY 401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420
| | | | |
Db 1720 GACGTGACCAAGGAGGATCAAGTTTCATCAATGAACGGGCTTATGGCCGACCCCATG 1779
QY 421 LysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGlnThrPhe 440
| | | | |
Db 1780 AAGGTGTACAAAGACCGCCAGGTCTGAACATGTGAGAGACAGAGAGAGAACCTTC 1839
QY 441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
| | | | |
Db 1840 CCGAGAAAGTTATGATGATGATCCCAAGAACTTTGGCTGATGTCATCTCTGAGAGAG 1899
QY 461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480
| | | | |
Db 1900 AAGACAGTGGCTGAGTGGCTCTCTATTACTACAGTCAAGAAAGATGAAACTATTAAG 1959
QY 481 SerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGln---GlnGlnGlnGln 499
| | | | |
Db 1960 AGCTGTGTGAGACGAGGCTATCGCGCCCGCAAGAGCACAGNACAGACGACGAG 2019
QY 500 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 519
| | | | |
Db 2020 CAGCAGACAGACAG 2079

[illegible]

Db	3108	GA	CGCGGAGGCGGCTGAGCGCA	CGGCGGAGGGGGGCTCAAGGCAAGAAAGAGGAGCG	3167
Qy	899	GI	YSerGIYArgIaLThrThraLalYSerSerGIYAlaProGlnIAspSerAspSer	Ser	918
Db	3168	GG	AGAGCGGCAAGGAGCCACCA	CAGCCAGCAAGAGCTCGGGCGCCCCCAGGACAGGACATCCACGT	3222
Qy	919	Al	athrCYsSerAlaAspGluVala	AspGluValaGluGlyAspLYsAspAArgLeuLeu	938
Db	3228	GCT	ACTGCTGAGTGAAGACGAGGTGATGAGGG	CGAGGGCGGGGACAAAGACCGGCTGGCTG	3287
Qy	939	Ser	ProArgProSerSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGln		958
Db	3288	TC	CCCAAGGCGCCAGCGCTCTCCAC	CCCCGACTGGCGAGCCCCCGGGCCAAAGCTCCACCCGAG	3347
Qy	959	LY	SProLeuAspLeuLYsGlnLeuLYsGlnArgAlaAlaIleProProIleGlnVal		978
Db	3348	AA	GCCTGAGACTTGAAGAGCTGAAGAGGAGGAGCGGTGGCAATCCCCCATC	---GTC	3400
Qy	979	Thr	LYsValAlaGluProProArgGluAspAlaAlaProThrLYsProAlaProPro-Al		998
Db	3405	AC	CAAAAGTGCATGAGCCCCCGGGAGGAGCGAGTCCACCAAGCCAGCTCCCCCAAGC		3466
Qy	998	A	ProProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSe		1018
Db	3465	CC	CAACCGGACCGGCAAAACCTGAGCGCCGAGAGGAGCCCGCTCAGCAGCTCGGACAG		3524
Qy	1018	P	roArgGlyLYsSerArgSerProAlaProProAlaAspLYsGlu	-----	1033
Db	3525	CC	CGCGGGGACAGAGCAAGAGCCGGGACCCCCCGGCGACAAAGAGGACAGAAAGCTGT		3584
Qy	1034	-----	-AlaPheAlaAlaGluValaGlnLYsLeuProGlyAspProProCYsTrpTh		1050
Db	3585	GT	TTCTCCCAACCTTTCGAGCGCCAGAGCCCAAGACTGCTGGGACCCCCCTTGCTGAC		3644
Qy	1050	T	SerGlyLeuProPheProValProProArgGluValIleLYsAlaSerProAlaAlaP		1070
Db	3645	TY	CCGCGCTGCGCTTCCCGCGTCCCGCCCGCGAGGTGATCAAGGCTTCCCGCATGCCCC		3704
Qy	1070	CA	SPProSerAlaPheSerTrpAlaProProGlyAlaProLeuProLeuGlyLeuAlaAs		1090
Db	3705	GA	CCCCCTCAACCTTCTCTCAAGCTCCACCTGAGTACCACATGGCCCCCTGAGCTTCATGA		3764
Qy	1090	P	ThraLArgProValLeuProArgProProThrIleSerAspProProLeuAlaIleSe		1110
Db	3765	CA	CTGCCCGGGCGCTTCCGCGCGGCCCAACCATCTCCAAACCGGCTCCCTCATCTC		3824
Qy	1110	T	SerAlaLYsAlaPheSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSe		1130
Db	3825	CT	GTGCCAAGACCCCGAGCGTCTTCGAGAGGCAATAGTGCATCTCCCAAGAAATGTC		3884
Qy	1130	r	ValGlnLeuAlaValProTYrSerGlyAlaAlaLYsAlaProValAlaGlyProValAla		1150
Db	3885	GG	TCCAGCTCCACCTCCGTACTGAGCAAGCCAAAGCCCCGGTGGGCGCTGTCCACAT		3944
Qy	1150	CG	IleuProLeuProMetAspProLYsLYsLeuAlaProPheSerGlyValLYsGlnG		1170
Db	3945	GG	GGGCTGCGCTTGGCCATGACCCCCAAAAAGCTGGACCTTTCAGCGGAGTGAAGCAGGA		4004
Qy	1170	u	GlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThraAlaG		1190
Db	4005	GC	ACGTGTCGCCCAAGGGGCGAGGCTGGGACCCGGAAGAGCTGGGGGTGCCACAGCCCA		4064
Qy	1190	n	GluAlaSerValLeuArgGlyThraAlaLeuGlySerValProGlyLYsLeuTrpLY		1210
Db	4065	GG	AGAGCGCTCGCTGTGAGAGGGAAGCTCTGGGCTCACTTCCGGCGGAAGCATCACAA		4124
Qy	1210	sg	IlePProSerThraArgValProSerAspSerAlaIleThrTYrArgLYsLeuTh		1230
Db	4125	AG	GCATTCCTCCAGCAACGGGAGCGCTTGGAGACGGCCATCATACCGCGGCTTCATCAC		4184
Qy	1230	r	HisGlyThrProAlaAspValLeuTYrLYsGlyThrIleThraArgIleIleGlyGluAs		1250

Db 4185 CCACGGCAGCCAGCTGACCTCTCTGTAACAAGGCAACCATACCAAGGATCATCGGCGAGA 4244
Qy 1250 pSerProSerArgLeuAspArgGlyValArgGluAspSerLeuProGlyGlyHisValIleTyr 1270
Db 4245 CAGCCCGAGTCCGTTTGACCGCGGCGGAGAGACAGCTGCCAAGGGCCACGTCATCTTA 4304
Qy 1270 rGluGlyValGlyHisValLeuSerTyrGluGlyMetSerValThrGlnCysSe 1290
Db 4305 CCAAGGCAAGAGGGCCAGTCTTGCTCTATGAGGGGTGGCATGTCTGTAGACCAAGTGTCTC 4364
Qy 1290 rTysGluAspArgValArgSerSerSerGlyProProHisGluThrAlaAlaProGlyArgHis 1310
Db 4365 CAGAGAGGACGGGAGAAAGAGCTCAGAGACCCCCCTGAGAGCGGCCGCCCAAGCGCAC 4424
Qy 1310 rTyrAspMetMetGluGlyArgValGlyValArgAlaIleSerSerAlaSerIleGluGlyLe 1330
Db 4425 CTATGACATGATGAGAGGCGCGCTGGGCAAGGCACTCTCTCAGCCACGATCGAAGTCT 4484
Qy 1330 uMetGlyArgAlaIleProProGluArgHisSerProHisHisLeuGlyGluHisHis 1350
Db 4485 CATGGGCGGTGCATCTCCGCGGAGGACAGACCCCAACACCTCAAGAGAGACCA 4544
Qy 1350 rTleuArgArgGluAlaLeuLeuLeuGlyArgGluGlyThrProProProProProSe 1390
Db 4545 CATCGCGGGGTCCATACACAAGAGATCCCTCGGTCTTACGTGAGAGCAACAAGAGACTTA 4604
Qy 1370 rLeuArgArgGluAlaLeuLeuLeuGlyArgGluGlyThrProProProProProSe 1390
Db 4605 CTTCGCTCCGGAGGCCAAGCTCTTAAAGCGGAGGCAACCTCCGCCCCACCGGCTCTC 4664
Qy 1390 rArgAspLeuThrGluAlaTyrTyrLeuThrGlnAlaLeuGlyProLeuGlyLeuLeuProAl 1410
Db 4665 ACGGGACTGACCGAGGCTTACAAAGACGAGGCGCTCGGCGCCCTGAAGCTGAAGCGGCTC 4724
Qy 1410 rHisGluGlyLeuValAlaThrValGlyGluAlaGlyArgSerIleHisGluIleProAr 1430
Db 4725 CATGAGGGCTGTGTGGCCACGGTGAAGAGCGGGCGCTCTCATCATGATGATTCGCGG 4784
Qy 1430 gGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuGlyGlySe 1450
Db 4785 CAGAGAGCTGCGGCAACGCCCGAGCTGCCCTCGGCGCGCGGCTCTCAAGAGGGCTCTC 4844
Qy 1450 rIleThrGlnGlyThrProLeuGlyTyrAspThrGlyAlaSerThrThrGlySerLysIy 1470
Db 4845 CATCAGCGAGGACACCCGCTCAAGTACACACACCGCGCGCTCCACCATGCTGCCCAAAA 4904
Qy 1470 rHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProle 1490
Db 4905 GACCGACGTACGCTCTCTATCGGCAGCGCGCGGAGCGTCCACCGGTGACCGGCT 4964
Qy 1490 uAspValMetAlaAspAlaArgAlaLeuGluAlaGlyAlaCysTyrGluGlySerLeuGlySe 1510
Db 4965 GGAATGATGTGGCCGACGCCCGGGCACTGGAACTGTGCTGTAGAGAGAGCTTGAAG 5024
Qy 1510 rArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleVal 1530
Db 5025 CCGGCGAGGAGACCGCAGAGAGCTCGGGGGGCTCATTTGGCGCGGCGGCCCGGTCAATTGT 5084
Qy 1530 lProGluuGlyValProArgGlnSerProLeuThrTyrGluAspHisGlyValAlaProh 1550
Db 5085 GCGTGTGCTGTGGAAACCGCGGAGAGCCCTTAACCTATGAGAGACCAAGGGCACCTTT 5144
Qy 1550 rAlaGlyHisLeuProArgGlySerProValThrMetArgGluProThrProArgLeuG 1570
Db 5145 TCGCGGCGACCTCCACGAGTTCGCGCGGTGACCAAGCGGAGCCACCGCGGCGCTTGA 5204
Qy 1570 nGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProAr 1590
Db 5205 GAGGGGACGCTTTCTGTCAGCAAGCATCCAGACCGAAAGCTGACGTCGACGCTCG 5264
Qy 1590 gGluIleAlaLysSerProHisSerThrValProGluHisHisProHisProIleSerPr 1610
Db 5265 TGAGATCGCCAAAGTCCCGGACAGACCGTGCACGAGCACCAACCCCACTTCGCTC 5324

Qy 1610 oTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProle 1630
Db 5325 CTATAGGACACTGCTTCGGGGCGTAGTGGCGTGAACCTGTATCCGACCACTACCTCCCT 5384
Qy 1630 uAlaPheAspProThrSerIleProArgGlyValLeuLeuAspAlaAlaAlaAlaTyrTy 1650
Db 5385 GGCCTTCGACCCCACTTCATACCCCGGCAATCCCTCTGGAGGAGCGGCTGGCTACTTA 5444
Qy 1650 rLeuProArgHisLeuAlaProAspProThrTyrProHisLeuTyrProProTyrLeuI 1670
Db 5445 CTTCGCCGACACCTGGCCCCCAACCCCACTACCGGACCTGTACCCACCTTACTTAT 5504
Qy 1670 rArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAspAspTyrI 1690
Db 5505 CCGCGGCTACCCCGACACCGCGCGCTGGAGAACGGGACCAATCATCATGACTACTACT 5564
Qy 1690 rHisSerGlnGlnMetHisHisAspThrAlaPheAlaMetAlaGlnArgAlaAspMetLe 1710
Db 5565 CACTTCGACGAGATGACACACACGCGCACCGCCATGGGCCACAGCTGATATGTCT 5624
Qy 1710 uArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgG 1730
Db 5625 GAGGGGCTCTGCGCCCGGAGATCTCGGTGGCATCTAAGCTGGGGGTCCCGAGG 5684
Qy 1730 yIleIleAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyTh 1750
Db 5685 CATCATCGACCTGTCCAAAGTCCACACTGCTGTGTGTCGCCCCGACACAGGCA 5744
Qy 1750 rProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSe 1770
Db 5745 CCAAGCACCGGCATGAGACCGGCTTGCTTACTCCCAACCGGCGCCACCTTCACGAG 5804
Qy 1770 rArgHisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrTh 1790
Db 5805 CCGCCACAGCACTCCCACTCTCCCAAGAGTCCAAACACTTGACAAACCAACCAAC 5864
Qy 1790 rHisSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgAlaArgGluAr 1810
Db 5865 CAGTCTCTGCTCGAGCGGAGGAGAGACCGGATGACAGCGGAGCCGGGATCGGAGGG 5924
Qy 1810 gGluLysSerIleLeuThrSerThrThrThrValGluHisAlaProIleTyrArgProG 1830
Db 5925 GGAAGAGTCACTCTCATGCTCCACACAGCGGTGAGAGCGGACCATGTGAGACTGG 5984
Qy 1830 yThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerAr 1850
Db 5985 TACAGAGCAGACAGGCGAGCAGCGGAGCGGCGGGGTGGGGGAGCAGCAGCGCG 6044
Qy 1850 gProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAl 1870
Db 6045 CCGCGCTCTCCACTCCATGCCACACAGCACTCGGCCCATCTCCCTCGGAGCCAGGATGC 6104
Qy 1870 rLeuGlnGlnArgProSerValLeuHisAspThrGlyMetLysGlyIleIleThrAlaVal 1890
Db 6105 CCTCCAGAGAAACCAAGTGTCTTCAACAACAGGCAAGAGGATATATACACGCTGT 6164
Qy 1890 lGluProSerLysProThrValLeu-----ArgSerThrSerThrSerSerProVa 1907
Db 6165 GAGAGCCACAGCCCAAGCTCTGTAGGTGGGCGAGGTCACTCCACTCCGACCCGT 6224
Qy 1907 lArgProAlaAlaThrPheProProAlaThrHisCysPProLeuGlyGlyThrLeuAspGl 1927
Db 6225 TCGCCCGGTGACACTTCCCACTGCGACCACTGCGGAGGGGGGAGCCCTCATGTG 6284
Qy 1927 yValTyrProThrLeuMetGluProValLeuLeuProGlyGluAlaProArgValAlaAr 1947
Db 6285 GGCTTACCTTACCTCATGAGACCGGTCTTGCTGGCCAAAGAGGCCCGGGGTCCCG 6344
Qy 1947 gProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSe 1967
Db 6345 GCCAGAGCGGCCCGAGCAGACACCGGCGCATGCTTCTCGGCAAGCCCGGCGGCTC 6404

QY 1967 rGlyLeuGluProAlaSerSerProSerLyseGlySerGluProArgProLeuValProPr 1987
 Db 6405 CGGGCTTGAGACCGCGCTCTCTCCCGACGAAGGCTCGAGCGCGCGCTTAGTGGCTCC 6464
 QY 1987 oValSerGlyHisAlaThrIleAlaArgThrProAlaLysAlaLeuAlaProHisHisAla 2007
 Db 6465 TGCTCTGGCCAGCCACCACTGGCCCGCTCGACCAAGAACCTCCACCTCACCAAGC 6524
 QY 2007 aSerProAspProProAlaProProAlaSerAlaSerAspProHisArgGlyLeuThrG 2027
 Db 6525 CAGCCCGAGACCGCGCGCGCTGCTCGCTCGGACCGGACCGGAGAAAGACTCA 6584
 QY 2027 nSerLyseProPhaSerIleGlnGluLeuGluLeuArgSerLeuGlyTrpHisGlySerSe 2047
 Db 6585 AAGTAAACCTTTCTTCATCCAGGAACCTGAACTCCGTTCTCTGGGTTACCAAGCGCACAG 6644
 QY 2047 rTySerSerProGluGlyValGluProValSerProValSerSerProSerLeuThrHisAs 2067
 Db 6645 CTACAGCCCGAAGGGGTGAGCCCGCTCAGCCCTGTAAGCTCACCCAGTCTGACCCACGA 6704
 QY 2067 pLyseGlyLeuProLyseHisLeuGluGluLeuAspLyseSerHisLeuGluGluLeuAs 2087
 Db 6705 CAAGGGGCTCCCAAGACCTCGAAGAGCTCGACAAAGGCCACTGAGGGGAGGCTCGG 6764
 QY 2087 gProLyseGlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuAr 2107
 Db 6765 GCCCAAGCAGACGAGCCCGCTGAAAGCTTGGCGGGGAGCGCGCCACTCCACACACTCGG 6824
 QY 2107 gProLeuProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLy 2127
 Db 6825 GCGCGTGCCTGAGAGCAGCCCTGCTCAGCCCGCTGCTCAGACCCCGCAGGGGTCAA 6884
 QY 2127 sGlyHisGlnArgValAlaThrLeuAlaGlnHisIleSerGlyValIleThrGlnAspTy 2147
 Db 6885 AGGTATCACAGCGGGGTGTCTACCTTGCGCCAGCATAGTAGGTCTATCACAGGACTA 6944
 QY 2147 rThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTySerPhePr 2167
 Db 6945 CACCCGGACACCAACAGCAGAGCTCAGCGACCCCTGCGCCCTTACTACTCTTCCC 7004
 QY 2167 oGlyAlaSerCysProValLeuAlaLeuAlaGlnArgProProSerAspLeuTyLeuProPr 2187
 Db 7005 TGGGGCAGCTGCCCTGCTCTGAGACCTCCGCCGCCACCAAGTGAACCTTCACTCCGCC 7064
 QY 2187 oProAspHisGlyAlaProAlaArgLyseProHisIleSerGlyGlyLysArgSerPr 2207
 Db 7065 CCGGAGCATAGTGTCCCGCGCGCTGCTCCCGCCACAGGAGGGGAGAGTCTCC 7124
 QY 2207 oGluProAsnLysThrSerValLeuGlyGlyGlyLysArgGlyIleGluProValSerPr 2227
 Db 7125 AGAGCCAAACAAACAGCTCGGTCTTGGGTGTGTGAGGAGACGATATTGAACCTGTCTCCC 7184
 QY 2227 oProGluGlyMetThrGluProGlyHisIleSerArgSerAlaValTyProLeuLeuTyAr 2247
 Db 7185 ACCGAGGGGATGACGAGCAGCAGGCACTCCGAGAGCTGTACCGCTGTCTGACCG 7244
 QY 2247 gAspGlyGluGlnThrGluProSerArgMetGlySerTySerProGlyValAsnHisSerG 2267
 Db 7245 GATTTGGGAACAACGAGGCCAGCAGATGGGCTCCAACTCTCCAGGCAACACAGCCA 7304
 QY 2267 nProProAlaPhePheSerLyseLeuThrGlySerAsnSerAlaMetValLysSerLyse 2287
 Db 7305 GCGGCAAGCTTCTTCAAGCAAGCTGACCGAGCAACTCGGCATGTGTCAAGTCAAGAA 7364
 QY 2287 sGlnGluLysLeuLysLeuLeuAlaThrHisAsnArgAsnGlnProGluTyArgHisIleSe 2307
 Db 7365 GCAAGAGATCAACAAAGAGTGAACACCCCAACCGGAATGAGACTGAAATACATATCAG 7424
 QY 2307 rGlnProGlyThrGluLysLeuPheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTy 2327
 Db 7425 CCGAGCTTGGGACGGAAGATCTTCAATATGCGCCGCAATACCGGAACAGGCTTATGACCTA 7484
 QY 2327 rArgSerGlnAlaValGlnGluHisIleSerThrAsnMetGlyLeuGluAlaIleIleAr 2347

Db 7485 TAGAAGCCAGCGGTGACGAGAACATGCACACCAACATGGGGCTGAGGCCATTAATTAG 7544
 QY 2347 glyAlaAlaLeuMetGlyLysTyThrArgIntTpgLysLeuSerProProLeuSerAlaAsnAl 2367
 Db 7545 AAAGGACCTATGATGGTAATATATGACAGTGGAAAGATCTCCCGCGCTCAGCGCAATGC 7604
 QY 2367 aPheAsnProLeuAlaAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAs 2387
 Db 7605 TTTTAACCTCTGAATGCCAGTCCAGCTCGCCCGCTGCTATGCTCAATACCGTGTGA 7664
 QY 2387 pGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaLysValSerG 2407
 Db 7665 CGGACGAGTGAACACACACTCACTCGCAGGTGGCGCGGGAAGGCCAAGGCTCTCGG 7724
 QY 2407 YArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerLysAspArgPr 2427
 Db 7725 CAGACCCAGCAGCGAAGAAACCAAGTCCCGCGCGCGCTGCACTTGGGAGACCGGCC 7784
 QY 2427 oProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAs 2447
 Db 7785 ACCCTCTGTCTCTCACTGCTGCTCGAAGGAGACTGCAACCGCGGACGCGCTCACCA 7844
 QY 2447 nArgValITTPGluAspArgProSerSerAlaGlySerThrProPheProTyArgnProle 2467
 Db 7845 CCGCGTGTGGAGAGACAGGCGCTGTCGCGAGGTTCCAGCAGCATTCCTTACAAACCCCT 7904
 QY 2467 uIleMetArgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeuProAlaG 2487
 Db 7905 GATCARGCGGCTCAGAGCGGGGTCTCATGGCTTCCACCCCGCGGCTCCCGCGGG 7964
 QY 2487 ySerGlyProLeuAlaGlyProHisHisAlaITPAspGluGluProLysProLeuLeuCy 2507
 Db 7965 CAGCGGCGCTCTGCTGCGGCCCAACAGCCTTGGAGAGAGAACCCAAAGCACTGCTCTG 8024
 QY 2507 sSerGlnTyGluThrLeuSerAspSerGlu 2517
 Db 8025 CTCGAGTACGAGACACTCCGACAGCGAG 8055
 RESULT 4
 AF113001
 LOCUS
 DEFINITION
 Mus musculus silencing mediator of retinoic acid and thyroid hormone receptor alpha mRNA, complete cds.
 ACCESSION
 AF113001
 VERSION
 AF113001.1 GI:4454547
 KEYWORDS
 Mus musculus (house mouse)
 SOURCE
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 8388)
 Ordentlich,P., Downes,M., Xie,W., Genin,A., Spigner,N.B. and Evans,R.M.
 Unique forms of human and mouse nuclear receptor corepressor SMRT
 Proc. Natl. Acad. Sci. U.S.A. 96 (6), 2639-2644 (1999)
 99178941
 10077563
 2 (bases 1 to 8388)
 Downes,M.R., Ordentlich,P. and Evans,R.M.
 Direct Substitution
 Submitted (11-DEC-1998) Gene Expression Laboratory, The Salk
 Institute for Biological Studies, 10010 North Torrey Pines Road, La
 Jolla, CA 92037, USA
 FEATURES
 source
 1..8388
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /tissue_type="spleen; brain"
 626..8047
 /function="transcriptional co-repressor"
 /note="mSMART alpha"

```
/codon_start=1
/product="silencing mediator of retinoic acid and thyroid
hormone receptor alpha"
/protein_id="AAD20944.1"
/db_xref="GI:4454548"
/translation="MSGGQPVAAQVMTAAEPYPPHGISVYQVIAASHNDVGLAEYOH
HPRDITSHLSFSSIIQPOARRPSLSSEPPGSEBSOEHLARESKRTFLDELKQDIER
TESKPRLELPDLTLRPSPLATQPSGSEBDLTQDRSLAGLELPEVSPSPHAPDEL
ELAPRLSKKEELIQNMDRVREITWVEQOISKLKKQOQLEBAAPPEKPEVAPP
IESKRSILQIYDENRKAFAAHRIIEGLGQVELPLVQIOPSDITQVYENIKIQAM
RKLLIYFKRRHARKOMERQCORQDOLMAEMEKVEEIEENNPDRQVYENIKIOM
KOPFLRKORELOERMOSRVQORGSLSMARSEHVEIIDGLSEONLEKROYE
AVIPMLYDADQRIKFTMMGLMDPMKYVYKDRQVTTMWSGDERPTPEKWKQPKA
FGLIASFERKTVAECVLYYLTKKNENTKSLVRSYRRGKSQOQOQOQOQOQOQOQ
AASSOBEKEKEKEKEADBEERKODABEKEBELSEKTDTSQEDDHEKEAASGKR
TANSQGRKRGIITRSMANENHETATPOOSSELSAMENNESRMTKEEMETAKGIL
EHRMWSAIARWGSKTVGOCQNFYNYKRONLDEILIOHKLKMEKENARKKKKT
PAASEETAAPPAADEMEWASGANEELAEAAQASQANEVPRVECSGPAAVN
NSDTSBVSPPSAEMKDTGPRGTGTEALPAATQPPVPEEPAPAAAPSPVPDASG
PSPSPSHLPHRLMTWTKMKPRLLQVLRQMPKSRSLRRSRSEPEEPSEASEE
PSPSVKSHKEETEERBEDKAGTEAIEVTSAPLKEVAGKAATVKGSSGATQDS
DESATCSADEVEDPEEGDKRLSPSPSLTPAGDPASTSFOKPLDKQLKORAAI
PIQVTKVSEPREDTVPKVPVPVPPPOHLOQEGDVQOQSGSPRGSRSPVPAE
KKAERAPPAAPRTGPKLPTRPPMSSGLPPPIPREVITKTSPPHADSASVTPG
HELPLGHDASAPVLPRPISNPPELISAKHPGULERLQALISQOQMSVQLRVHSE
HAKPMGFLTWGLPLAVDKLUGTALGATSISITKGLPSTRAADSPSTRSITHEPT
ADVLVKGTSISRTVGEDSPRLDRLAREDTLPKHGVLYEGKGHVLSEEGMSVQSCKE
DORSSSGPHEHETAPKRTYDMMGKRGVTVTSASIEGLMGRALPEQHSMLKOHIR
GSITQGIPRSVEAOEDYLREAKLKEKGTGPPPPDLTETVYKRPDLDEPKL
KTHHGVNAVTEKAGRSIHEIPREELRTPELAPRBEKSIITGQPELTKDQSGPS
TCTKGDVASTIGSPRRPPALHPIDIMADALBACREBSIKSKSSGTSAGASGIT
RQAPVVPPELGRKROSPLYEDHAPFTHLPGSGVTTRETPRLQESLSSKASQ
DKLTLSTPREIAKSPHSTVPEHHPHPISEYELHRCVTVNDVLRGHIPLAPFTSLPR
GIPLEAAAAVYLPRLAPSPYPLPYLLIRGYDITALENRQTIINDVITSQOMH
HNAASMAQADMLRGLSPRESSLALNVAAPRGIDLSQVHLPLYVPEPTGTPATA
IDRLAYLTPAAPPSRSHSSPSLSPGCPHILAKPTATSEERREBEREDRSITLST
TTVEHAPTRPCTGSSGAGSSRPASHTHGHSIPSPRIODALQORPVIHNTSMKV
VTSVEGPTVLMARASTSTSSVPVREAPPPATHCPIQDLEGGVPTIMEVPLKE
TSRVARPERAVDAGHAFLLTKPRPPASPSKSSPRSLAPSPSSIAATPAAVLN
APHHASDPAPPTASADLREKTSQKPSIOELIELSLRSLAGSPGVPISIVS
SPSLTNDKLSLEBELKESHLGELRHQOPGMKLSAAHALPHLRPEPOSSSP
LIOTAPGIGHORVYLAOHISEVITQDTRHHPQOLSLYPSIPFGACCPVLDL
RRPESDLYLPDPDHGTPRAGSPHSEGGKSPSPSKTSVUGSEDAIEPSPREGNTER
GHARSTAYPLLYRDGQGRPMGLSPGNTSQPTEFSKLTSSNSMNVSKQELNKK
LWTHRNNEBEYNIQOPGTETI FMMPAITGALTCRQAAVOGHAISTPGGALIRKML
GYKDOWEPEPLGANAFLNLSASALTPAAMPITTAGDSHATLSPGCGAKYSGR
PSSRKKSPAPGIASGDRPSPSVSVEGDCNRRTPLTRVWEDRPSAGSNPPYNP
LIMRLQAGVMASSPPPGGLAAGSGLAPGHANDBEKPLLCQOYETLSSE"
711..1187
/misc_feature
/note="Region: deleted in SMRT beta isoform encoded by
GenBank Accession Number AF113002"
```

```
ORIGIN
Alignment Scores:
Pred. No.: 2,74e-215 Length: 8388
Score: 11067.00 Matches: 2160
Percent Similarity: 88.20% Conservative: 89
Best Local Similarity: 84.71% Mismatches: 192
Query Match: 83.75% Indels: 110
DB: 10 Gaps: 29
US-09-522-753-5 (1-2517) x AF113001 (1-8388)
QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTPAArgAlaThrGluProArgTYrPro 20
Db 626 ATCTGAGATCCACAGCCTGTGGACACAGCATGCGGGCTCTGACCCCGCTACCCA 685
QY 21 ProHisSerLeuSerTYrProValGlnIleAlaArgThrHisThiAspValGlyLeuLeu 40
Db 686 CCCCATGATCTCTTCAACCCGCTGCAAGATAGCCCGCTCCACACGAGCGGGCTGCTT 745
QY 41 GluTYrGlnHisSerArgAspTYrAlaSerHisLeuSerProGlySerIleIleGln 60
```

```
Db 746 GAGTACCAACACACACCCCGTGACTTACACCTTACACCTGACCCGGTTCCATCATCCAG 805
QY 61 ProGlnArgArgProSerLeuLeuSerGlnProGlyAsnGluArgSerGln 80
Db 806 CCACAGAGGAGGCGGCCCTCACTGCTGTACAGATTCCAGCTCGGAGTGAAAGCGTCTCAG 865
QY 81 GluLeuHisIleLeuArgProGlnSerHisSerTYrLeuProGlnGluGlyArgSerGlnMet 100
Db 866 GAGCTCCACCTGCGCCCTGAGTCCCGACGCTTCTGCTGAGCTGGGCAAGCCCGACATA 925
QY 101 GluPheIleGlnSerIleArgProArgLeuGlnLeuLeuProAspProLeuLeuArgPro 120
Db 926 GAATTCAACGAGACGACAGCGCCCGCTGAGACTACTACCGGATACCTGCTGGCCCA 985
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGlnAspLeuThrIleAspArgSer 140
Db 986 TCACCCCTGCTGCGCACCTGGACGAGGTGGGTCTGAAGACTTAACCAAGACCGCTAGC 1045
QY 141 LeuThrGlyIleLeuGlnProValSerProProSerProHisThiAspProGlnLeu 160
Db 1046 CTGGCAGGCGMACCTGAGCTGTGTCACTCCACAGTCCCGACGCTGACCCCTGAGCTA 1105
QY 161 GluLeuValProProArgLeuSerIleGlnIleGlnAsnMetAspArgValAsp 180
Db 1106 GAGCTGGCGCCCATCTCGACTGTCCAAAGAGAGAGCTGATCAGAAATGAGACCGCGTGAC 1165
QY 181 ArgGluIleThiMetValGlnGlnGlnIleSerIleValLeuIleGlnIleGlnIle 200
Db 1166 CCGTAGATACCATGATGTAGAGAGACGAGATCTCCAAAGCTGMAAGAGACGAAACGTTG 1225
QY 201 GluGlnGluAlaIleValProProGlnProGlnIleValSerProProProIleGln 220
Db 1226 GAGGAGAGAGCGCCGACAGCGCCGACGAGAGCTGTGTGGCACACCCATAGAA 1285
QY 221 SerIleHisIleArgSerLeuValGlnIleIleTYrAspGluAsnArgIleValIle 240
Db 1286 TCMAAGCACCGAAGCTGTGTCCAGATCATCTACATGAGAAACCGAAGAAACCGAAGCC 1345
QY 241 AlaHisArgIleLeuGlnIleGlnIleValProGlnIleValGluLeuProLeuTYrAsnGlnPro 260
Db 1346 GCACACCGGATCTTGAAGAGGCTGGGGGCCCGAGGTGAGACTCTCTGTACCAACGACCG 1405
QY 261 SerAspThrArgGlnTYrHisGlnLeuAsnIleValIleAsnGlnAlaMetArgIleValLeu 280
Db 1406 TCTGACACACGCGCAGTACCATAAACATCAAAATTAACACAGCGATGGGAAAGAAAGCTG 1465
QY 281 IleLeuTYrPheIleArgArgAsnHisIleAlaArgIleValGlnIleValSerPheCysGln 300
Db 1466 ATCTTGTTACTTTTAAGCGAGGAACACACGCGCCGACAGCATGGGAAACAGCGCTTCTCCAG 1525
QY 301 ArgTYrAspGlnLeuMetGlnIleLeuGlnIleValIleValIleGlnArgIleGlnLeuAsnPro 320
Db 1526 CGCTATGACCACTCATGAGTGGAGCGGTGGGAGAAAGTGAAGCGCATAGAAACAATCCG 1585
QY 321 ArgArgArgAlaIleValGlnSerIleValArgGluTYrTYrGlnIleValGlnPheProGlnIle 340
Db 1586 CGAAGAGGCGCAAGAGAGCAAGGTGAAGAGGTACTAGAAACAGATGTCCCGGAGATC 1645
QY 341 ArgIleValArgIleLeuGlnIleValArgMetGlnIleAspArgValIleGlnArgIleSerGly 360
Db 1646 CCGAAGCAGCGGAGCTGACGAGACCGCATGACAGACGAGGTGGCCACGCTGGCACTGGG 1705
QY 361 LeuSerMetSerAlaIleArgSerGlnHisGlnIleValSerGlnIleIleAspGlyLeuSer 380
Db 1706 CTCTCCATGTCGGCTGCCCGCAGTGAAGTGAAGTTCCTGAGATCATTAATGAGCTTGTCT 1765
QY 381 GluGlnGlnIleAsnLeuGlnIleValGlnIleMetArgGlnIleAlaValIleProProMetLeuTYr 400
Db 1766 GAGCAGGAGAACCTGAGAGAGAGAGATGCGCAGCTGGCGCTGATTCGCCGCAATGTATC 1825
QY 401 AspAlaAspGlnIleArgIleIleValPheIleAsnMetAsnGlyLeuMetAlaAspPromet 420
Db 1826 GACGCGGACCAAGCAGAGATCAAGTTCTCAATCAATGAATGAATGATGATGACCCCATG 1885
```


OY	421	LYSVALLTYLVASPRAGLIVAlmeclanmetTPSERGIUGLIVLysGLIthrPhe	440
Db	1886	AAGGCTCAAGAGACCGTCAGGTTTCCAACTAGTGAGCGAGCAAGAGGACACCTTTC	1945
OY	441	ARGVLVSPHMEGLNIhIaProLYaAnPheGLIleuLIeAlaSerPheLengluYrg	460
Db	1946	CGTGGAAGTTTATACGACCCCTTAAGAACTTGGCTGATTTGGCTCAATCTCGAAGGA	2005
OY	461	LYSTHVALAAGLUCYValLeuTYTYTYTYLeuThrLYbLYaAnGLIuAnTYrLYbS	480
Db	2006	AAGAGCGTCGCTGAATGTGTCTCTTATTACTACCTGACCAAGAAATGAATAATTCACAG	2065
OY	481	SerLeuVALARGASerTYrARGrgrGLYbLYSserGLINGLINGLINGLIn	500
Db	2066	AGCTTGATGAGGCGGAGCTATCGCGCGTGGCAAGACGACAGACAGACAGAGCA	2125
OY	501	GLINGLINGLInGLINGLInGLINGLInProMetProARGSerSerGLIUGLILyS	520
Db	2126	CAACGACGACGACGACAG-----ATGGACGAGAGGACAGAGAGAGAG	2173
OY	521	ASPGLYbGLYbGLYbGLYbGLIuAAGLILySGIUGLILyProGLIValGLIuAn	540
Db	2174	GAGAGAAAGCAAGAGAGAGAGAGAGGCCCAAGAGAGAAAGAGACAGATGCGGAGAC	2233
OY	541	AsPLYbGLIuAsPLeuLeuLYbGLIuLYrThrAsPThrSerGLIuAsPAsnAspGLIu	560
Db	2234	GAGAAAGCAAGACTCAGCAGAGAGAGACAGACGACACTTCGCGGAGGACACACATAG	2293
OY	561	LYbGLIuAVALAAserLYbGLIYARGLYrThrAlaAnSerGLINGLYARGrGLY	580
Db	2294	AAAGGGCGCGGCTCCAAAGGCGCGCAAACTGCCAAACAGCACAGCGCGCGCAAGGC	2353
OY	581	ARGIETRRASerMetAlaAnGLIuAlaAnSerGLIUGLILeThrProGLIn	600
Db	2354	CGTATCACCGCTCCATGACCAAGCGACCAACTGAGAGACAGCCACCCACAGCA	2413
OY	601	SerAlaGLIuLeuAlaSerMetGLIuLeuAnGLIuSerSerARGrTPThrGLIUGLImet	620
Db	2414	AGTTACAGAGCTGGCTTCATGGAAGTGAACGAGATTCTCGCTGAGCTGAAGAGAGATG	2473
OY	621	GLIthrVALbLYbGLYLeuLeuGLIuHIbGLYARGAnTPSERAlIeAlaArgMet	640
Db	2474	GAGACAGCAAAAGAAAGGCTCTCGGACATGGAGAACTGTGACGATTCGCCGATG	2533
OY	641	VALGISerLYSTHVALSerGLIbLYbAsnPheTYrPheAnTYrLYbLYARGIn	660
Db	2534	GTGGGCTCCAAAGCCGTGTCCACAGTGAAGACTTCACTTCACTCAACAAAGAGGAG	2593
OY	661	AsnLeuAspGLIuLIeLeuGLINGLInHIbLYbLYeLYbMetGLIuLYbGLIARGAsnAlaArg	680
Db	2594	AACTCGACGAAATCTCTTACAGACACAAAGCTTAAAGATGGAAGAGAGAGAAAGCTTCGG	2653
OY	681	ArgLYbLYbLYbLYbAlaProAlaAlaAlaSerGLIuAlaAlaPheProProValAla	700
Db	2654	AGGAAGAAAGAAAGAACCCGACGTGGCGGCGAGCGAGACAGCTTCCACCTGCCGCT	2713
OY	701	GLIuAspGLIuImetGLIuAlaSerGLIYAlaSerGLYbAnGLIUGLImetValGLIUGLIn	720
Db	2714	GAGACCAAGAGATGAGAGACATACGCGCCCAAGTCCATGAGAAAGCTGGCGAGAGAG	2773
OY	721	AlaGLIuAlaLeuHIbAlaSerGLIYAsnGLIuValProArg---GLIYGLIYbSerGLIPro	739
Db	2774	GCAAGACCTCACAGGCTCTTGGAATAGATTTCCAGAGTTGGGAGTGCAGTGGCCCA	2833
OY	740	AlaThrVALAsnAsnSerSerAspThrGLIuSerILEProSerProHIbThrGLIuAlaAla	759
Db	2834	GCTCGTCGCAACAACGCTCTGATACTGAGAGTGTCCATCCCGCGTTCAAGAGCCATG	2893
OY	760	LYbAspThrGLIuGLIAnbGLIProLYbProProAlaThr-----LeuGLYAlaAspGLY	777
Db	2894	AAGACACT-----GGGCTTAAACCACTGACCTGAAGCATTTGGCCGCTGCACCC	2944

QY	778	ProProProGlyProProThThProProArgrarhThSerArGAlaProIleGluProThr	797
Db	2945	CAGCCACCTGTTCTCT-----CCTCCAGAGAACCGGAGAGTACCCCTGCTGAGGCCCTCC	2998
QY	798	ProAlaSerGluAlaThrcGlyAlaProThrProProProAlaProProSerProSerAla	817
Db	2999	CCAGTCCCTGATGCCAGTGGGCCACCATCTCCCAAGACCTTC-----CCATCACTGGCCCA	3055
QY	818	ProProProValProLySGluGluLySGluGluGluThrAlaAlaAlaProProVal	837
Db	3055	CCCCGGCTACGTGTGACAGAGATGACACAGAGCCCGGGGCTCCAGCTCCCAACGA	3114
QY	838	GluGluGlyGluGluGluGluProProAlaAlaGluGluLeuAlaValAspThrGlyLe	857
Db	3115	GAGGATGCCAAGAGAGACAGAGTGTGAGGCCGAGGAG-----ATGATGTGGGAA	3165
QY	858	Ala-GluGluPro-----VallySerGluGlyThrGlu	868
Db	3166	GCAGAGGAGCCCGAGGCTCTGAGAGACCCCGAGAGTGTAAAGATGTGACACAGGA	3225
QY	868	GluGluAlaGluGluGlyProAla---LySGlyLyAspAlaGluAlaAlaGluAlaThrAl	887
Db	3226	GGAAGCCGAGAGAGAGCTTGAAGACAAAGCCAGAGGCACAGAGCCATGTAAACTGTCTC	3286
QY	887	AGluGlyAlaLeuLyAlaGluLyAlaGluLyGlySerGlyAlaThrAlaThr--Al	906
Db	3286	TGAGGCACACCTTAAAGTGAGAG-----GAGGCTGTATGACAGGACAGCTGTGACCAAGG	3339
QY	906	AlaSerSerSerGlyAlaProGluAlaAspSerSerAlaThrGlySerAlaAspGluVal	926
Db	3340	TTCCAGCTCAGTGTCCACCCAGGACAGAGACTTCAAGTCCACCTGCAGTGCCTGATGAGT	3399
QY	926	LAAspGluAlaGluGlyGlyAspLyAspAsnArgLeuLeuSerProArgProSerLeuLeuTh	946
Db	3400	GGAAGAACCCGAAAGAGGTGACAAAGGCGAGGCTGTGTACCAAGGCCAGCTCTCTAC	3455
QY	946	rProThrGlyAspProArgAlaAsnAlaSerProGluLyAspLeuAspLeuLyGluLe	966
Db	3460	CCGGGCTGTGAGATCCCGGGCGAGTACTGCGCCAGAAAGCCGCTGAGCTGAAGACGT	3519
QY	966	uLySGluArgAlaAlaAlaAlaLeuProProIleGluValThrLySValHSGluProProAr	986
Db	3520	GAACAGAGAGACACCGGCATCCCCCTATCCAGGTACACAAAGGTCCATGAGCCCCCGG	3579
QY	986	GGAlaAspAlaAlaProThrLyAspProAlaProProAlaProProProGluAsnLeuGlu	1006
Db	3580	GGAAGACACAGTACCCCAAGGCCAGTCCCCCTGTCCAGGTACACCAAGGTCCATGAGCTTCA	3639
QY	1006	nProGluSerAspAlaProGluGluGluGlySerSerProArgGlyLySLeuSerArgSerPr	1026
Db	3640	GCCAGAGGCTGACGTGTCTCAGCAGTGGGAGGAAAGTCCACGTGGAAAGTCCGACAGCC	3699
QY	1026	AlaProProAlaAspLySGlu-----AlaPheAlaAlaGlu	1038
Db	3700	AGTGCCTCTGTCGAGAAAGGAGCAGAGAAACCCGCAATTCCTTCCGCTTCCCAACTGA	3759
QY	1038	uAlaGluLyLeuProGlyAspProProGlySerThrSerGlyLeuProPheProValPr	1058
Db	3760	GGGCCCCAAGGTACGACTGAGGCCCAAGCGGTGTACTGGGCTGTGCTTCCCACTCC	3819
QY	1058	oProArgGluValIleLyAlaSerProHAlaAlaProAspProSerAlaPheSerTyAla	1078
Db	3820	TCCAGCGGAGGTGTACAACTTCCCAACGCGCGTGACCCCTGTGCTTCTCTAC	3879
QY	1078	aProProGlyAlaProLeuProLeuGlyLeuHAlaAspThrAlaArgProValLeuProAr	1098
Db	3880	ACCCCGCGTACCCCGCTGTGCGGCTTCCAGATAGTGGCCGCGCGCTGCGACAG	3939
QY	1098	gProProThrIleSerAsnProProProLeuIleSerSerAlaLyHAlaProSerValLe	1118
Db	3940	TCCCCC---ATCTCTAACCCCAACCCCTCATCTCTGTGCAAGATATCCGGCGTACT	3996
QY	1118	uGluArgGluIleGlyAlaIleSer---GluGlyMetSerValGluLeuHAlaProTy	1137

```

Db 3997 TAGAGAGGAGCTGGGTGCTATCCACAGAGGGAGTGCATCCAGCTTCGATGCTCTCA 4056
Qy 1137 rsergluhtsalalyalaProValGlyProValThrMetGlyLeuProLeuProMetAs 1157
Db 4057 CTGAGGCAATGGCCAGAGCCCAATGGGCTCTCACCAATGGGGCTCCCTTCCTGGGA 4116
Qy 1157 pProlyslvleuAlaProPheserGlyValylsGlnGlnLeuSerProArgGlyG1 1177
Db 4117 CCTTAAGAAAGCTG----- 4129
Qy 1177 naIaglyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgG1 1197
Db 4130 -----GG 4131
Qy 1197 yThrAlaLeuGlySerValProGlyGlySerIleThrIysGlyIleProSerThrArgVa 1217
Db 4132 GACAGCACTGGGCTCCGCCACCACTGGAAAGCATCAACCAAGGCTCCCAATGCCGGGC 4191
Qy 1217 lProSerAspSerAlaIleThrTyArgGlySerIleThrHisGlyThrProAlaAspVa 1237
Db 4192 TGCAGAGGGCCCC-----AGCTACAGAGGCTCTATACCCACGGCACCCCGCAACCT 4245
Qy 1237 lLeuTyArgGlyThrIleThrArgGlyIleGlyGluAspSerProSerArgLeuAspAr 1257
Db 4246 CCTTACAGAGGATCATCAGACAGATCGTCGGTGAGAGACAGCCCAAGTCGCTTGACCG 4305
Qy 1257 gGlyArgGlyuAspSerLeuProlyGlyHisValIleTyArgGlyValylsGlyHisVa 1277
Db 4306 GGCACAGAGAGAACCCCTGCCCAAGGGCCATGTCATCTAGAGGGCAAGAAAGGCCACGT 4365
Qy 1277 lLeuSerTyArgGlyGlyIleMetSerValThrGlnCysSerGlyGluAspGlyArgSerSe 1297
Db 4366 CCTATCTTATGAAGTGTATGTCCGTGCACAGTCTCTTAAGAGAGATGGAAGAGCGAG 4425
Qy 1297 rSerGlyProProHisGlyThrAlaAlaProIysArgThrTyArgMetMetGlyValr 1317
Db 4426 CTCGGGCCCCACCCCATGAGACTGCCCCCTTAACACACTATGACATGATGAGGGGCGG 4485
Qy 1317 gValGlyArgAlaIleSerSerAlaSerIleGlyGlyLeuMetGlyArgAlaIleProPr 1337
Db 4486 TGTAGGCAAGAGCTGTGATCTCCAGCCAGCATAGAGGATCTATAGGGCCGGCCATC--CC 4542
Qy 1337 oGluArgHisSerProHisIleLeuIysGlyGlnHisIleArgGlySerIleThrG1 1357
Db 4543 TAGAGCAACAGCCCC--CACTCAAGAGAGCATCACTCGAGGCTCCATCAGCA 4599
Qy 1357 nGlyIleProArgSerTyValGlyAlaGlnGlyuAspTyTLeuArgArgGlyAlaIysIe 1377
Db 4600 AGGCAATCCCGAGGCTCTATGTGAGGGCGGAGGACTACTTACGGCGGAGGCCAAGCT 4659
Qy 1377 uLeuIysArgGlyGlyThrProProProProProSerArgAspLeuThrGluAlaTy 1397
Db 4660 CTTGAAGCCAGAAAGGACACACCCACCCACCACTCGGAGCTGATGAGACCTA 4719
Qy 1397 rIysThrGln-----AlaLeuGlyProLeuIysLeuIysProAlaHisGlnGlyIe 1414
Db 4720 CAAGCCCGGCCCCCTGGAGCCCTGTGGTCCCTGAAGCTGAAGCCGACACAGAGGCTGT 4779
Qy 1414 uValAlaThrValIysGlyAlaGlyArgSerIleHisGlyIleProArgGlyGlyLeuArg 1434
Db 4780 GGTAGCAACCTGTGAAGAGGCGGGCGCTTATCTATGAGATCCGAGAGAGAGCTGGG 4839
Qy 1434 gHisThrProGluLeuProLeuAlaProArgProIysGlyGlySerIleThrGlnG1 1454
Db 4840 CCGCAACCTGAGCTACCTCCCTGGCACCGCTCTGAAGAGAGGTTCATCCATCCAGAGG 4899
Qy 1454 yThrProLeuIysTyArgThrGlyAlaSerThrThrGlySerTyIysHisAspValAr 1474
Db 4900 CACCCCACTCAATAGACTCTGGGGCACTCCATCGGACCAAGAAACACACACTGGG 4959
Qy 1474 gSerLeuIleGlySerProGlyValrGThrPheProProValHisProLeuAspValMetAl 1494

```

```

Db 4960 CTCATCATTCGGACGCCCGGCGCTTTCCTGCTCCCTGCACCCGCTGGACATATAGGC 5019
Qy 1494 aAspAlaArgAlaLeuGluArgAlaCysTyArgGlyGlyLeuIysSerArgProGlyTh 1514
Db 5020 TTAACGCCCGGGGACTGGAGCGGTGCTGTATGAAGAGTCTGMAAGGCCGCTCAGGAGC 5079
Qy 1514 rAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGlyLeuG1 1534
Db 5080 CAGCAGTGTGTGAGGGGCTCCATCAACGTGGGGCTTCAGTGTGTGCTGAACTGGG 5139
Qy 1534 yIysProArgGlnSerProLeuThrTyArgIleuAspHisGlyAlaProPheAlaGlyHisIe 1554
Db 5140 CAAGCCAGGCAAGCCCACTGACTTACAGAACACAGGGGCACTTACACATGCACCT 5199
Qy 1554 uProArgGlySerProValThrMetArgGlyuProThrProArgLeuGlnGlySerIe 1574
Db 5200 GCCAGTGGCTCCCTGTATCCACAGAGGAGCCAGGCCCTTCAAGAAAGGACCT 5259
Qy 1574 uSerSerTyAlaSerGlnAspArgIysLeuThrSerThrProArgGlyIleAlaIy 1594
Db 5260 CTTATCCAGCAAGGGCTCCAGAGCCGGAAAGCTGACATCTACACCCCGGAGATGCCCA 5319
Qy 1594 sSerProHisSerThrValProGluHisIleAspHisProIleSerProTyGlnHisIe 1614
Db 5320 GTCCCAACACAGACTGTGCCGAGACACACCTTACCCATCTCCCTATGAGCACTT 5379
Qy 1614 uLeuArgGlyValSerGlyValaAspLeuTyArgSerHisIleProLeuAlaPheAspPr 1634
Db 5380 GCTCCGGGGCGTGAAGTGTGTGACTGTATCGGTGTACATCCATGTGGCTTGACCC 5439
Qy 1634 oThrSerIleProArgGlyIleProLeuAsp---AlaAlaAlaAlaTyTyrLeuProAr 1653
Db 5440 CACTTCATACCCCAAGGATCCCTGGAACAGAGCGGAGCCCTACTACCTGCCCG 5499
Qy 1653 gHisLeuAlaProAspProThrTyProHisIleuTyArgProTyTLeuIleArgGlyTy 1673
Db 5500 GCATTTGGCCCCAGCCCCCATCTACACACCTGACACCTTACCTTATCGGGCTA 5559
Qy 1673 rProAspThrAlaAlaLeuGluAspArgGlnThrIleIleAspTyTyrIleThrSerG1 1693
Db 5560 CCTTACACGGGGCGCTTGAGAACCGGCACATCATCATATGACTATCATCCTCGCA 5619
Qy 1693 nGlnMetHisIleAspThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgIyIe 1713
Db 5620 GCAGATGACACCAACAGCTGCTCCGCGCATGCGCCAGCTGTGACATCTGAGGGGT 5679
Qy 1713 uSerProArgGlyuSerSerLeuAlaLeuAspTyAlaAlaGlyProArgGlyIleIleAs 1733
Db 5680 GTACCCGAGAGTCTGCTGCTGCTCATATATGCGCTGGCCCAAGAGCATTTATCA 5739
Qy 1733 pLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAlaTh 1753
Db 5740 CTTGTCCCAAGTGCACACACTGCGGTGTGTGACCAACAGCCAGGACCCCTGCCAC 5799
Qy 1753 rAlaMetAspArgLeuAlaTyTyrLeuProThrAlaProGlnProPheSerSerArgHisSe 1773
Db 5800 CCGCATCGACCCCTTGCTTACTCCCTCCACTCGGCCGCCACCTTCAGACGCGCCACAG 5859
Qy 1773 rSerSerProLeuSerProGlyGlyProThrHisIleuThrIysProThrThrThrSerSe 1793
Db 5860 TAGCTTACCGGTGTCCCAAGAGGCCCACTCACTAGCTTAACCAATGCCACATCTTC 5919
Qy 1793 rSerGluArgGluArgAspArgGlyuArgAspArgGlyuArgAspArgGlyuArgIyIysSe 1813
Db 5920 ATCCAGACCGGAACGGGAACGTGAGCGGAAGAGC-----AAGTC 5961
Qy 1813 rIleLeuThrSerThrThrValGlnHisAlaProIleTyrArgProGlyThrGlnG1 1833
Db 5962 CATCTCAGCTTACACTAAGTGAAGATCACCATCTGAGACCTGTGATGAGAGCA 6021
Qy 1833 nSerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgProAlaSe 1853
Db 6022 GAGCAGCGGGGCT-----GGGGCGACGAGCGGCCCGCTC 6057

```


ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 8544)
Park, E.-J., Schreier, D. J., Yang, M., Li, H., Li, L. and Chen, J. D.
SMRte, a silencing mediator for retinoid and thyroid hormone
receptor corepressor
Proc. Natl. Acad. Sci. U.S.A. 96 (7), 3519-3524 (1999)
JOURNAL 99199215
PUBMED 10097068
REFERENCE 2 (bases 1 to 8544)
AUTHORS Chen, J. D.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-1999) Pharmacology and Molecular Toxicology,
University of Massachusetts Medical School, 55 Lake Avenue North,
Worcester, MA 01655, USA
FEATURES
source
1. .8544
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
1_8544
/gene="Smrte"
160..7548
/gene="Smrte"
/function="transcriptional corepressor"
/note="SMRte"
/codon_start=1
/product="silencing mediator of retinoic acid and thyroid
hormone receptor extended isoform"
/protein_id="AAD2972.1"
/db_xref="GI:4559296"
/translation="MSGSTQPVATQWRAAPRYPHGISVYQVQIARSHTDVGLAEYOH
HRDYSHTLSPGSIITOPORRPSLSEPPGSRSGOELRLRPSRLRFLKIDPDL
TSSKPRLELPDITLRPSPLATGPPSSSDITKRSIAGKEPSPSPRNADDEL
ELASPRLKEELIQNRLDVRDEITWEEQISLKKKQOQLEBAKPEPEKPSVP
PIESKRSIYQIYIDENRKAENAHRIELGLQVQLPYNQSIDRQYHENTIKINQA
MKKILYFKRNHARKWQRCORIDLEAMEKKEVRIENPPRAKESKVRBY
EKOPFIRKORELOERMOSRGVORGSLNSARSHRYSEIIDGLSEBENLEKMO
LAVIRHVVRDQRIKFTINMGIMDDPMVYKDRQYTNMSEDEPRTKMPKHPN
FGILASFLEKRTVAECVLYYLTKKENTKSLVRSYRRKGSQOQOQOQOQOQOM
AASSQSEKEKEKEADKEEKQDAENEKEBELSKERTDITSENDKEVAASKK
TANSQGRKGRITRSMANENHETAPQOOSSELAEMNEMSSRLKEENARKKKT
EHRGNWAIARWGSKTVSQCKNFYNYKRNQLDEILQOKIKMEKERNARKKKT
PAAASETAPPAADEEMBAEGASANEELAEAEASOGSGNEVVRGEGSGPAVN
NSQDTSVSPSEAEATKQDQPKPTGTEALPAATOPVPPPEEPAADSPADSG
PPSPSPSPAPAPATVDKQAPAPAPAPQTDAKGQKSEAEITVGRPEPEABE
PPSPVSDHKEETEEPEDEKAGTEALITVSEAPLVEEASAAVTKSSSGATODS
DSATCSADBEDEEGDKGRLSPRPSLITPAGDRASTSPQKPLDLQKQRAAI
PPIVTVHPREDVTPPKVPVPVPPPTQHLQPEGDVSOQSGSPGKRSVPAPAK
BAEKAPPAFPTGEOGYRLSPHAGRLSPHREVIKSTRADPLFSYTPGRLPL
GLHDSARPLRPPISNPRLISSAKHNGEROLGALISGAKVQVLRVPSSEAKMG
PLTMEPLAVDPKKTALAPPVESAPRASQYPCGRQLOLYHPRPADVLYGCT
ISRIVEDESPSLDKARBDTLPRGHVITYEKGHIVLSYEGMSVSCSDKDGSSGP
PHETAPKRTYDMGRVGTIVTSASIEGLMGRAIPBOHSPIKEDHNRGSLTQGP
RSVEAQDYIYREAEKILKREGTPPPREDLTETKPRPLDLKPKRHEVAV
ATYKAGRGIHEIYREBLRRTPELPAPRPLKESITQGPRLYDGAESTKXKHEV
BSITSGRPRPALPIDIMADARALEACVRESLSKSGTSGAGGSIIRGAPVVP
ELGKPRQSLTYEDHCAPTTSHLPRGSPVTTRETPRLQSGLSKASQDRKLTTP
RETAQPSHSTVEPHNPISVYHLRGVGVLYNGHILPADPSTIPGILPLKAA
AAVYLRHLAPSPYPLHYPLIRGYPDPALENQOTINDYITSQOMHNHNSAMA
QRADMLRGLSPRESSALNYAGPRGIDLSQVPHLPVLPVPPGPATIDILAVLP
TAPPPSRHSSSPSPGPTHLAKPATATSEERERERESKISITSTTYENAPI
WRCTGSSGAGGSSRPSHTQHSPISPTODALQORBPVLNMTKMGVNSVEGT
PTYLASTSTSSVPRAATPPRATPCGGLTBEVYPTIMEPULIPRETSHVAPREPR
VODGHAFLTPAREPASSPSSSEPSLSLSPSSHTALIKRTAKSLAPHASPDG
PTASDLHREKTKSPESIOELRLSGYHSGAGYPDVEYSPVSPSLTDKGLS
KPLELEKSHLEGLRHKQGPMLKSAEANHLPLRLPSPQSSPLDQTAIGLGH
QRVTLAHOHISEVITQDYTHHPOOLSGPLPAPLYSPGASCVLDRRPSDLKLP
PDGTPARSGPHEGGRSPESKTSVLSKSDAITSPPSEMTBGRHARSVYPL
YRDEGEGEPMSKSPONTSQAPATPSTKTESANYSKKQGIKKKLTNHNNEBYL
NIGQPTETLPMPAITGAGLMTCRSQAVQOEHNASTNMGLEAIIRKALMGWEPPPP

ORIGIN		LCANAFNPINLASASLPAAAMPITTDGSDHAIJSPGGGKAKVSGRPSRKAASAP GLASDRPSSVSHSEKCNRTPTNKKVWEDRPSASSTFPYTNPLTRLOAGVMA SPPPGLAAGSGPLAGPHHAWDEBKPLCSQYETLSDSE"	
Alignment Scores:			
Pred. No.:	1.5e-210	Length:	8544
Score:	10832.50	Matches:	2149
Percent Similarity:	87.94%	Conservative:	90
Best Local Similarity:	84.41%	Mismatches:	197
Query Match:	81.97%	Indels:	117
DB:	10	Gaps:	33
US-09-522-753-5 (1-2517) x AFI25671 (1-8544)			
QY	1	MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro	20
DB	160	ATGTCAAGATCCACACAGCCCTGTGCACAGACATGGCGGGCTGTGAGCCCGCTACCCA	219
QY	21	ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu	40
DB	220	CCCATGGCATCTCCACCCGGTGCAGATAGCCGGTCCACACGATGGGGCTGCTT	279
QY	41	GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln	60
DB	280	GAGTACCAACACACACCCCGTGACTACCTACACCTTCACCCGTTCCATCATCCAG	339
QY	61	ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyValangLuarGserGln	80
DB	340	CCACAGAGAGAGCGGCGCTCAGCTGCTCAGAGTTCCACCTGGAGTGAAAGCGTCTCAG	399
QY	81	GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet	100
DB	400	GAGCTCCACCTCGCGCCTAGTCCCGCACGATGCTCTCCGACGCTGGGCAAGCCCGACATA	459
QY	101	GluPheIleGluSerTyrArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro	120
DB	460	GAATTACCGAGAGCAAGCGCCCGCTGAGACTTACCCGATACCTCTGCTGCCCA	519
QY	121	SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer	140
DB	520	TACACCTCTGCTGCGCACCTGGCAGCGGAGTGGGTGAAAGACTTACCAAGAGCCTGAGC	579
QY	141	LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeu	160
DB	580	CTGGCAGGCAACCTGAGCGTGTCTACCTCCACGATCCCGCAGCCTGAGCTA	639
QY	161	GluLeuValProProArgLeuSerLysGluLeuIleGlnAsn--MetAspArgVal	179
DB	640	GAGCTGGCGCCATCTTGACTGTCCAAAGAGACTGATCCAGAACGATTTGACCCGCTG	699
QY	180	AspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGln	199
DB	700	GACCGGAGATCAACCATGTAGAGCAGCAGATCTCCAACTGAAGAAAGCAGCAACAG	759
QY	200	LeuGluGluGlnAlaAlaLysProProGluProGluLysProValSerProProIle	219
DB	760	TTGGAGAGAGAGCGCCCGCAAGCGCCGGAACCGGAACCTGTGTGCGCACCCATA	819
QY	220	GluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGlu	239
DB	820	GAATCAAGACACGGAAGCTGGTGCAGATCATCTACAGTAGAAGAACCGAAGAACCGAA	879
QY	240	AlaAlaHisArgIleLeuGluGluGlyLeuGlyProGlnValGluLeuProLeuTyrArgGln	259
DB	880	GCGGACACACGATCTTACAGAGGCTGGGGCCCAAGTGAAGTGTCTGTACCAACAG	939
QY	260	ProSerAspThrArgGlnTyrHisGluAsnIleLysIleLeuGlnAlaMetLysLys	279
DB	940	CCGTCTGACACACGCGCAGATCATGTAACCAATCAATCAACGCGAGTCCGGAAGAG	999
QY	280	LeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCys	299

Dh 1000 CTGATCTTTGATCTTTAAAGCGGAGAAACACGCGCGCAAGACGTGGGAAACAGCGCTTCTGC 1059
Qy 300 GlnArgTyrAspGlnLeuMetGluAlaLeuGluIlybSylValGluArgIleGluAsnAsp 319
Dh 1060 CAGGCGCTATGACCAAGCTCATGTGAGCGCGTGGAGAAAGAAAGTAAAGCGCATAGAAACAAT 1119
Qy 320 ProArgArgArgAlaIlybGluSerIlybValArgGluTyrTyrGluIlybGlnPheProGlu 339
Dh 1120 CCGGAAAGAGGGCGCAAGAGACAGGATGAGGAGTACTACGAGAAACAGTCTCCCGAG 1179
Qy 340 IleArgIlybGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgIlySer 359
Dh 1180 ATCCGCAAGACGCGGAGCTGCAGAGAGCATGACAGACAGGCTGGCGCACCGTGGCGCT 1239
Qy 360 GlyLeuSerMetSerAlaAlaArgSerGluIlybGluValSerGluIleIleAspGluLeu 379
Dh 1240 GGGCTCTCCATGTGGCGCTGCCGCAATGACATGAGTTTCTGAGATCATTTAGTGGCTTG 1299
Qy 380 SerGluGlnGluAsnLeuGluIlybGlnMetArgGlnLeuAlaValIleProProMetLeu 399
Dh 1300 TCTGACGAGGAGAACTGGAGAACAGATGCGCACGCTGCCGTGATCC--GCCATGTTG 1357
Qy 400 TyrAspAlaAspGlnGlnArgIleIlybPheIleAsnMetAsnGlyLeuMetAlaAspPro 419
Dh 1358 TACGACCC--GACCGACAGAGATCAAGTTCATCAACATGAATGACTCATGATGATGACCCC 1416
Qy 420 MetIlybValTyrIlybAspArgGluValIleMetAsnMetTyrSerGluGlnIlybGluThr 439
Dh 1417 ATGAAAGTCTTACAAAGACCGTCAAGTTTACAAATGTGAGCGACGAGAAAGGAGACCC 1476
Qy 440 PheArgGluIlybPheMetGlnIleAspIlybAsnAsnGlyLeuIleAlaSerPheLeuGlu 459
Dh 1477 TTCGCTGAGAAGTTTATGACCACTTAAAGACTTTGGCTGATTTGCTCTACTCTGGAG 1536
Qy 460 ArgIlybThrValAlaGluCysValLeuTyrTyrTyrLeuThrIlybIlybAsnGluAsnTyr 479
Dh 1537 AGAAAGCGGTGCTGAGTGTCTCTTACTTACTGACCAAGAAATGAAATTAATTAAC 1596
Qy 480 LysSerLeuValArgArgSerTyrArgArgArgIlybSerGlnGlnGlnGlnGln 499
Dh 1597 AAGAGCTTGTGAGGCGGAGCTTATGCGCGCGTGGCAAGAGCCGACGACGACGACGACG 1656
Qy 500 Gln 519
Dh 1657 CAACAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1704
Qy 520 LysAspGluIlybGluIlybGluIlybGluAlaGluIlybGluGluIlybProGluValGlu 539
Dh 1705 AAGGAGGAGAAAGGAGAAAGGAGAAAGGAGAAAGGAGAAAGGAGAAAGGAGAAAGGAG 1764
Qy 540 AsnAspIlybGluAspLeuLeuIlybGluIlybThrAspAspThrSerGlyIlybAspAsnAsp 559
Dh 1765 AAGGAGGAGAAAGGAGAAAGGAGAAAGGAGAAAGGAGAAAGGAGAAAGGAGAAAGGAG 1824
Qy 560 GluIlybGluAlaValAlaSerIlybGlyArgIlybThrAlaAsnSerGlnGlyArgArgIlyb 579
Dh 1825 GAGAAAGAGCGCGTGGCTTCCAAAGCGCCCAAACTCCCAACGACCAAGCGCGCGCAA 1884
Qy 580 GlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGln 599
Dh 1885 GGGCGTATCACGCGCTCATGCGCCAAAGAGCCCAATAGAGAGACAGCAGCCCGCAG 1944
Qy 600 GlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerArgTyrProGlnGluGlu 619
Dh 1945 CAAGGTCAGAGCTGGCTTCCATGAGAGATGAACGAGGTTCTGCTGACTGAGAAAGAG 2004
Qy 620 MetGluThrAlaIlybIlybGlyLeuLeuGluIlybGlyArgAsnTyrSerAlaIleAlaArg 639
Dh 2005 ATGAGAGACAGCAAGAAAGCGCTTCTGAAACATGGAGAGAACTGTGACGCAATGGCCGC 2064
Qy 640 MetValGlySerIlybThrValSerGlnCysIlybAsnPheTyrPheAsnTyrIlybIlybArg 659
Dh 2065 ATGATGGGCTCCAAAGACCGGTGCTCCAGTGTAAAGAACTTCACTCAAGAAAGAGG 2124

Qy 660 GlnAsnLeuAspGluIleLeuGlnGlnIlybIlybLeuIlybMetGluIlybGluAlaArgAsnAla 679
Dh 2125 CAGAACTCGAGCAAGAAATCTTACACACACAAAGCTTAAAGATGAGAAAGAGAAAGAACCT 2184
Qy 680 ArgArgIlybIlybIlybValAlaProAlaAlaAlaSerGluGluAlaAlaPheProProVal 699
Dh 2185 CCGAGAGAAAGAAAGAAAGACCCCGAGCTGGCGGAGGAGAGACAGCCTTCCACCTGCC 2244
Qy 700 ValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGlu 719
Dh 2245 GCTGAGAGCAAGAGATGAGAAAGCATCAGCGCAAGTCCCAATGAGAAAGACTGCGGAG 2304
Qy 720 GluAlaGluAlaLeuIlybAlaSerGlyAsnGluValAlaProArg---GlyGluCysSerGly 738
Dh 2305 GAGGCAAGAGCTTCAAGCGCTCTGGAGAAAGAGTTCCCAAGATTTGGGAGTGCAGTGGC 2364
Qy 739 ProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerProAlaIlybThrGluAla 758
Dh 2365 CCAGCTGCTGTCAACAAACAGCTGTATGAGAGTGTCCCATCCCGCGTTCAAGAGCC 2424
Qy 759 AlaIlybAspThrGlyGlnAsnGlyProIlybProProAlaThr-----LeuGlyAlaAsp 776
Dh 2425 ACGAAAGACACT-----GGGCTTAAACCCCACTGCACTGAAGCATGCGCGCTGCC 2475
Qy 777 GlyProProProGluProProThrProProArgArgThrSerArgAlaProIleGluPro 796
Dh 2476 ACCGACGACCTGTCTCT-----CTTCCAGAAAGACCGGACAGCGCCCTGCTGAGCCC 2529
Qy 797 ThrProAlaSerGluAlaIlybThrGlyAlaProThrProProProAlaProProSerProSer 816
Dh 2530 TCCCACTGCTCCATGTCAGAGTGGCCACCAATCCCAAGAGCTTCC---CATCACTGTGCC 2586
Qy 817 AlaProProProValValIlybGluGluIlybGluGluGluGluGluGluGluGluGluGlu 836
Dh 2587 GCACCCCGGCTACTGTGAGCAAGAGTGAACAAGAACCCCGGCTGTCCAGCTCCAGC 2646
Qy 837 ValGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 856
Dh 2647 ACGAGAGATCCAAAGAGAGCAAGAGTGTGAGGCCGAGAG-----ATCGATGTGGGA 2697
Qy 857 LysAlaGluGluPro-----ValIlybSerGluCysThr 867
Dh 2698 AAGCAGAGAGAGCCGAGCGCTGAGAGAGCCCGGAGAGTGTAAAGATGACACACAG 2757
Qy 868 GluGluAlaGluGluGluProAla---LysGlyIlybAspAlaGluAlaAlaGluAlaThr 886
Dh 2758 GAGGAGACCGAGAGAGAGCCTGAAAGCAAGAGGCAAGAGGCAAGAGGCAATTGAAACTGTG 2817
Qy 887 AlaGluGluAlaLeuIlybAlaGluIlybIlybGluGluGluGluGluGluGluGluGluGlu 905
Dh 2818 TCTGAGCACCACTTAAAGGTGAG-----GAGGCTGTGAGCAAGGACGCTGTGACCAAG 2871
Qy 906 AlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAspGlu 925
Dh 2872 GGTTCAGCTCAAGTGCACCCAGAGACAGTCACTCCAGTGCACCTCCTCACTGCGGAG 2931
Qy 926 ValAspGluAlaGluGluGlyIlybIlybAsnArgLeuLeuSerProArgProSerLeuLeu 945
Dh 2932 GTGAGCAACCCCAAGAGAGGTGACAAAGGCGAGCTGTGTACCAAGAGCCAGCTTCTC 2991
Qy 946 ThrProThrGlyAspProArgAlaAlaAlaSerProGlnIlybProLeuAspLeuIlybGln 965
Dh 2992 ACCCGCGTGAATCCCGGCGCAAGTCACTTCCGCCCAAGAGCGCTGACCTGAAACAG 3051
Qy 966 LeuIlybGlnArgAlaAlaAlaIleProProIleGlnValThrIlybValIlybGluProPro 985
Dh 3052 CTGAGAGAGAGAGACCGGCAATCCCGCTATC---GTCAACAAAGGTCCATGAGCCCCC 3108
Qy 986 ArgGluAspAlaAlaProThrIlybProAlaProProAlaProProProProGlnAsnLeu 1005
Dh 3109 CCGAGAGAGACAGTAAACCCCAAGAGCAAGTTCCTCCCTGTGCTTCCACCAAGAGACCTTA 3168

Dh 5191 TCACCGGAGAGCTCTCGCTGGCCCTCATATTATCCGCTGGCCCAAGAGCATTTATCGAC 5250
Qy 1734 LeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAlaThr 1753
Dh 5251 CTGCGCAAGTGCACACCTGCTGCTGTGTGTCACCAAGGACCGACCTGCGCAC 5310
Qy 1754 AlawerAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSer 1773
Dh 5311 GCCATCGACCGCTTGGCTCACTCCACCTGCGGCCCACTTCAGAGCGCGCACAGT 5370
Qy 1774 SerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSer 1793
Dh 5371 AGCTCAGCGGTGCTCCCGAGAGGCCCATCTACTAAACCACTGCCACATCTTCA 5430
Qy 1794 SerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSer 1813
Dh 5431 TCGGAGCGGAGACGGAACTGAGCGGGAACGAGAC-----AAGTCC 5472
Qy 1814 IleuThrSerThrThrThrValGluHisAlaProIleTPArgProGlyThrGluGln 1833
Dh 5473 ATCCTCAGCTCACTACACTAGTGAAGCATGCACTCTGAGACCTGTGTAGAGGAGCAG 5532
Qy 1834 SerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgProAlaSer 1853
Dh 5533 AGCAGCGGGGCT-----GGGGGACAGCACCGCCCGCTCC 5568
Qy 1854 HisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGln 1873
Dh 5569 CACAC-----CACGACCTGCCCCATCTCCCCGAGACCGAGGCGCTTGACACAG 5622
Qy 1874 ArgProSerValLeuHisAsnThrGlyMetLysGlyLysIleThrAlaValGluProSer 1893
Dh 5623 AGGCCAGTGTGTCGACACAGAGCATGAGGGCGGTGTCACCTCGTGGAAACCGGCG 5682
Qy 1894 LysProThrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaThrPhe 1913
Dh 5683 ACCGCCACGGTCTGAGTGCACCTCCACTTCGCTCGCCGACGCTGACCATTC 5742
Qy 1914 ProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrMet 1933
Dh 5743 CCACCTGCCACCACTGCCCACTTGTGGACCTTTAGAGGGGTCTACCTACCTCATG 5802
Qy 1934 GluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProAla 1953
Dh 5803 GAGCCGCTCTTACCAAGAGACCTCGGGTGGCCGCGCCGAGCGCCCGGTG 5862
Qy 1954 AspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSer 1973
Dh 5863 GAGGGTGGCATCTTCTCCACCAACCCCGCGCGG-----GAGCCGCTCC 5913
Qy 1974 SerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThr 1993
Dh 5914 TCACCCAGCAAGAGCTCCGAGCCCGATCCCTAGACACCCCGAGCTTCAGCCACACAGCC 5973
Qy 1994 IleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProAla 2013
Dh 5974 ATCGCCCGAGCCCGAGAAAGAGCTTGCACCCACATGCAATGCCAGCCGCGCGGG 6033
Qy 2014 ProProAlaSerLysSerAspProHisArgGluLysThrGlnSerLysProPheSerIle 2033
Dh 6034 ---CCCACTCGGCTCAGATCTGCACCGAGAAAGCTCAAGTAACCTTTTCATC 6090
Qy 2034 GlnGluLeuGluLeuArgSerLeuGlyTyrHis---GlySerSerLysSerProGly 2052
Dh 6091 CAGGAATTGGAACTCCGTTCTCTGGGTACCAAGTGAAGCTGCTACACCCCGAGTGG 6150
Qy 2053 ValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLys 2072
Dh 6151 GTGGAGCCCATCAGCCGGTGAAGCTCCCGAGCTGACCAAGCAAGGGGCTCTCCAA 6210
Qy 2073 HisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGly 2092
Dh 6211 CCTCTGAAGAGCTGAGAAAGACCACTTGGAGGGAGCTGCGGACCAAGCAGCAGGC 6270

Qy 2093 ProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGlySer 2112
Dh 6271 CCATGAAGCTCAGCGGAGAGGCTGCGCATCTCCCACTCTGCGGACACTGCCGAGAGC 6330
Qy 2113 GlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgVal 2132
Dh 6331 CAGCCCTCATCCAGCCCACTCTCCAGACTGCCCGAGGATCAAGAGGTCAACAGAGGTG 6390
Qy 2133 ValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspLysThrArgHisHisPro 2152
Dh 6391 GTCACTTGCTCAGCACTACAGAGTCAATTCAGAGACTACACCCGCGCACACCGC 6450
Qy 2153 GlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysPro 2172
Dh 6451 CAGCAGTCACTGAGCCCTTCCGCTCTCTACTCTTTCCCGAGCCAGCTGCTCC 6510
Qy 2173 ValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHisGlyAla 2192
Dh 6511 GTGCTGATCTTCGCGGCCCAAGCATGACTTACTCCCAACCCCGACATGGAC 6570
Qy 2193 ProAlaArgGlySerProHisSerGlyGlyLysArgSerProGluProAsnLysThr 2212
Dh 6571 CCAAGCCGGGATCCCGCCACAGTGAAGGGGCAAAAGTCCCGAACCACGCAAAACA 6630
Qy 2213 SerValLeuGlyGlyGlyLysAspGlyIleGluProValSerProProGlyLysMetThr 2232
Dh 6631 TCGGTCTGGGACAGCACTGAGAGTGCATTGAGCTGTGTGCCACCAAGGCGGATACT 6690
Qy 2233 GluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThr 2252
Dh 6691 GAGCCAGCATCTCGAGCGCTGTGTACCACTGCTGTATGAGACGGGAAACAGGCG 6750
Qy 2253 GluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePhe 2272
Dh 6751 GAGCCC---AGATGGGCTCTTAAGTCTCCAGGAAACCGACGCGCGCTTCTTC 6807
Qy 2273 SerLysLeuThrGlnSerAsnSerAlaMetValLysSerLysLysGlnLysIleAsnLys 2292
Dh 6808 AGTAAGTCACTAGAGCACTCCGCCATGTGTAAGTCGAAGAACGAGAGATCAACAG 6867
Qy 2293 LysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnLysSerGlnProGly 2312
Dh 6868 AAACCTCAACCCCAACCGGAACGAGCCAGATTAACAATTGGCCAGCTGGGACGAA 6927
Qy 2313 IlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaVal 2332
Dh 6928 ATCTTAACATGCCCGCCATCACTGAGCAGGCTTATGACCTGTAGAACCGAGCGGTG 6987
Qy 2333 GlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGly 2352
Dh 6988 CAAAGACGCGCACCAACATGGGGCTTAGAGCCATTATGAAGAGCACTCATGGGT 7047
Qy 2353 LysTyrAspGlnTProGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsn 2372
Dh 7048 AAATATGATCAGTGGAAAGGCCCGCGCTGGGGCCAAATGCTTTAACCCTGAAT 7107
Qy 2373 AlaSerLysSerLeuPro---AlaAlaMetProIleThrAlaAlaAspGlyArgSerAsp 2391
Dh 7108 GCCAGGCGCACTGCGCTGCTGTATGCCATTAACCACTGCTACGAGCAGAGTAC 7167
Qy 2392 HisThrLeuThrSerProGlyGlyGlyLysValAlaLysValSerGlyArgProSerSer 2411
Dh 7168 CAGCACTCACTCGCCAGTGGAGGTGGAAAGCCAAAGTCTCTGGCAGACTTACAGC 7227
Qy 2412 ArgLysAlaLysSerProAlaProGlyLysAlaSerGlyAspArgProProSerValSer 2431
Dh 7228 CGAAAGCCAAAGTGCACACACAGGCTAGGCTCGGAGACGAGACCCCTTGTCTCC 7287
Qy 2432 SerValHisSerGluGlyAspArgAsnArgArgThrProLeuThrArgAsnArgValTProGlu 2451
Dh 7288 TCAGTCACTCAAGAGGGGAGCTGCAATGCCCGAACCACCTCAACCAAGTGTGTGGAG 7347

Oy		2452	AsparagProSerSerIaGlySeThrPhoPheProTyraAnProLeuIlleMeArgLeu	2471
Dd		7348	GACGGCCCTCATCTGCAGGGTCCACGCCATTCCTCCCAACCCTTGATTATAGGCTA	7407
Oy		2472	GlnIaGlyValMetAlaSerProProProProGlyLeuProIaGlySerGlyProleu	2491
Dd		7408	CAGGCAAGGTGCATAGCCTCCCCCGCCCACCTGGCTTGGGGAGAAGCGGGCCCTTA	7467
Oy		2492	AlaGlyProHisHlaIaTrpAspGluGluProLysProLeuLeuCyseGlnTyrGlu	2511
Dd		7468	GCTGTGCCCAACAAGCTGGATAGAGAGCCAAAGCACCTGCTGTTCACAGATAGAG	7527
Oy		2512	ThrIeuSerAspSerGlu	2517
Dd		7528	ACACTCTCGACAGCGAG	7545
RESULT 6				
AF113002	LOCUS	AF113002	7465 bp	mRNA linear ROD 20-MAR-1999
DEFINITION		Mus musculus silencing mediator of retinoic acid and thyroid hormone receptor beta mRNA, complete cds.		
ACCESSION		AF113002		
VERSION		AF113002.1	GI:4454549	
KEYWORDS				
SOURCE		Mus musculus (house mouse)		
ORGANISM		Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Musinae; Mus. 1 (bases 1 to 7465) Odentlich,P., Downes,M., Xie,W., Genin,A., Spinner,N.B. and Evans,R.M. Unique forms of human and mouse nuclear receptor corepressor SMRT Proc. Natl. Acad. Sci. U.S.A. 96 (6), 2639-2644 (1999)		
REFERENCE	TITLE	2 (bases 1 to 7465)		
AUTHORS	Downes,M.R., Odentlich,P. and Evans,R.M.			
JOURNAL	Direct Submission			
PUBMED	Submitted (11-DEC-1998) Gene Expression Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA			
FEATURES	source	Location/Qualifiers		
	1..7465	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/db_xref="taxon:10090"		
		/tissue_type="spleen; brain"		
		363..7124		
		/function="transcriptional co-repressor"		
		/note="SMRT beta; longer SMRT alpha isoform encoded by GenBank Accession Number AF113001"		
		/codon_start=1		
		/product="silencing mediator of retinoic acid and thyroid hormone receptor beta"		
		/protein_id="AAD20945.1"		
		/db_xref="GI:4454550"		
		/translation="MSGSTQVPAOTWRAADRYPHGISYEVJATASHTPLYNOPSDPIRQYHEHTKTNAMRKGLIKYKRNRNARKMEGRPCORXYDLMAMEKKVRIRMPRRRAVESKRETRYEKOFBIRKORELOERMOSRVGSGLSMSARSSEHSITIDGISEONLEKOMRQLAVTPMLVDADOQRKTFTNMGLMDPMKVYKDRQVTNMSBOERDTPEFKMOHKNFGLIASFLERKTVAECVLYYLTKNENYKSIVRSYRRKSOQOOOOOOOOOAMARSOEKEKEKEAEDEBKODANEKEELSKETDTSGEEDENDEKAVALSKRTANSQGRRKGRTRTEMANEAHHETVAPOOSELASMMNSMRKTEEMETAKKLILHGHRWSALARYGSKYTSOCKNPYNYKKRONDELIOOKIKMYEKERNARKKRTPALAAEEATFPAAADEMEMESGASANEELABEBEASOASGENYPRVECSGPAAVNNSDSDESVPSESEIKTQIPPTTELPAATOVPVPPPEEPAAAPAESVPVDASGPSPEPSHLPHPRLLTRMNKMKPYLLQIPQMFRSLRPBRSAEMEKEBEASEAKSPKSVSHDKETEPEEPDKAGTEALIVSEDAPLKVERASKALAVTKSSGSGATODSDSASTCADVEDNEPDEPKNGRLISPBELLTPAGBPASTSPORKLDLKORRAAIPIPIYTKNAHPREDPVYPKPVPVPVPPYPOHLOPEDVSQQSGSSLRGKSSVPVPAKEAKEKPAFPAFPTBEPKPLTEBPWRSSGLPPIPIPREVIKTSPHAGMSVFTYTPGHPIGLIGHDASARVLPRPISNPPLISAKHGVLERDLGASIQAGMSVOLRAPHSEHAKAPMGPLTMGLIPLAVLDPRKUOTALSGATSSTGTLGETPAADOC		

[illegible]

Db 467 ----- 467
 Qy 201 GluGluGluAlaAlaLysProProGluProGluLysProValSerProProIleGlu 220
 Db 467 ----- 467
 Qy 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAla 240
 Db 467 ----- 467
 Qy 241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPro 260
 Db 468 ----- 485
 Qy 261 SerAspThrArgGlnTyrHisGluAsnIleLysIleLeuGlnAlaMetArgLysLeu 280
 Db 486 TCTACACACCCCGATACATGAAAACTCAAAATTAACCGGCGATCGGAAAGAGCTG 545
 Qy 281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGln 300
 Db 546 ATCTTGACTTAAAGCGAGAACCAAGCGCAAGAGTGGAACAGCCCTTCTGCCAG 605
 Qy 301 ArgTyrArgGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnPro 320
 Db 606 CGCTATACACGCTCATGAGCGCTGGAGAAAGATGAGCCCATAGAAACATCCG 665
 Qy 321 ArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340
 Db 666 CGAAGAGGCGCCAGAGAGACAGACAGTAAAGAGTAACTACGAAACAGTTCCCGAAGTTC 725
 Qy 341 ArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGluGlnArgLysSerGly 360
 Db 726 CGCAAGCAGCGGAGCTGCGAGAGCGCATGCAAGCGAGGTGGCCAGCTGGAGAGTGG 785
 Qy 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
 Db 786 CTCCTCCATGTGCGCTGCCGAGTGAAGCATGAGGTTTCTGAGATCATTTGCTGTCT 845
 Qy 381 GluGlnGluAsnLeuGluLysGlnMetArgGlnLeuValIleIleProPheLeuTyr 400
 Db 846 GAGCAGAGAACTCGAGAAACAGATGCGCAGCTGCGCTGATCCGCGCATTTGTAC 905
 Qy 401 AspAlaAspGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspPheMet 420
 Db 906 GAGCGGACCAAGAGATCAAGTTTCATCAACATGATGACATCAATGATGATGCCATG 965
 Qy 421 LysValTyrLysAspArgGlnValMetAsnMetTyrSerGluGlnLysGluThrPhe 440
 Db 966 AAGGTCTAACAGGACCTGTGATCCAAACATGTGGAGCGAGCGAGAGGACACCTTC 1025
 Qy 441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
 Db 1026 CGTGAGAAATTATGACGACCCCTAAGAACTTTGGCTGTGATGCTTATTCCTGGAGAGA 1085
 Qy 461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480
 Db 1086 AAGACGCTGCTGATGTCTCTTACTTACTGACCAAGAAATGAAAAATTACAG 1145
 Qy 481 SerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGln 500
 Db 1146 AGCTTGTTGAGCGGAGCTATCGCGCGCTGGCAAGACCAAGCGAGCGAGCGAGCAA 1205
 Qy 501 GluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
 Db 1206 CAACAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1253
 Qy 521 AspGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLys 540
 Db 1254 GAG 1313
 Qy 541 AspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGlu 560
 Db 1314 GAGAAAGAAAGAACTCAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1373

Qy 561 LysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnIleArgArgLysGly 580
 Db 1374 AAGAGGCGCTGGGCTCTCAAAAGCCGCAAACTCCCAACAGCCAGGCGCGCAAGGC 1433
 Qy 581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGln 600
 Db 1434 CGTATCAGCGCTCCTACGCGCAACGAGCCCAACATGAGAGAGAGAGAGAGAGAGAGAG 1493
 Qy 601 SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTyrThrGluGluMet 620
 Db 1494 AGTTCAAGCTGCTTCCATGAGATGACGAGAGTTCTGCTGAGCTGAGAGAGAGATG 1553
 Qy 621 GluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTyrSerAlaIleAlaArgMet 640
 Db 1554 GAGACAGCAAGAAAGCCCTCTGGAACATGGAGAACTGTGATGCCATTTGCCGATG 1613
 Qy 641 ValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysArgGln 660
 Db 1614 GTGGGCTCCAAGACCGTGTCCAGTGAAGACTTCACTTCACTCAAGAAAGAGCGAG 1673
 Qy 661 AsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArg 680
 Db 1674 AACCTGACGAAATCTTTCAGACAGCACAGCTAAAGATGAGAGAGAGAGAGAGCTCG 1733
 Qy 681 ArgLysLysLysLysAlaProAlaAlaSerGluGluAlaIlePheProProValIle 700
 Db 1734 AGAAGAGAAAGAAAGACCCCACTGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1793
 Qy 701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGlu 720
 Db 1794 GAGAGCAAGAAAGATGGAAGCATCAGCGCAAGTGCATAGAGAGAGAGAGAGAGAGAG 1853
 Qy 721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArg--GlyGluCysSerGlyPro 739
 Db 1854 GCAGAGGCTCACAGGCTCTGGGAAATGAGGTTCCAGAGTTGGGAGTCAAGTGGCCA 1913
 Qy 740 AlaThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAla 759
 Db 1914 GCTGCTTCAAACAAAGCTGTGATGAGAGTGTCCATTCCTCCGCTTCAAGAGCCAGC 1973
 Qy 760 LysAspThrGlyGlnAsnGlyProLysProProAlaThr-----LeuGlyAlaAspGly 777
 Db 1974 AAGGACACT-----GGGCTTAAACCACTGGGACCTGAGCATTTGCCGCTGCAC 2024
 Qy 778 ProProProGlyProProThrProProArgArgThrSerArgAlaProIleGluProThr 797
 Db 2025 CAAGCAACTGTTCT-----CTCCAGAAAGAACCGGCACTAGCCCTGTGAGCCCTCC 2078
 Qy 798 ProAlaSerGluAlaThrGlyAlaProThrProProProAlaProProSerProSerAla 817
 Db 2079 CAGTCCCTGATCCAGTGGCCCAACATCCCAAGAGCTTC---CCATCACTTGCAGCA 2134
 Qy 818 ProProProValAlaProLysGluGluLysGluGluGluThrAlaAlaProProVal 837
 Db 2135 CCCCCTGCTACTGTGAGCAAGATGAACAAGAGCCCGGCTCTCCAGTCCCAAGACA 2194
 Qy 838 GluGluGlyGluGluGluLysProProAlaAlaGluGluLeuAlaValAspThrGlyLys 857
 Db 2195 GAGAGTGCCAAAGAGAGAGAACTGAGCGCAGAGAG-----ATCGATGGGGAAAA 2245
 Qy 858 Ala-GluGluPro-----ValLysSerGluCysThrGly 868
 Db 2246 GCCAG 2305
 Qy 868 uGluAlaGluGluGlyProAla---LysGlyLysAspAlaGluAlaAlaGluAlaThrAl 887
 Db 2306 GAGAGCCGAG 2365
 Qy 887 agLysGlyAlaLeuLysAlaGluLysLysGluGlySerGlyArgAlaThrThrAlaLys 907
 Db 2366 TGAGGACCACTTAAAGTGAAG---AAGGCTGTGACCAAGGAGAGAGAGAGAGAGGTTT 2422

QY 907 sSerSergIyAlaProGlnAspSerAspSerSerAlaThrCySerSerAlaAspGluValAs 927
DB 2423 CAGCTCAGGTGCACACGACGACGATGCTCCAGTGCACCTGACAGTGCAGGATGAGGTGA 2482
QY 927 pGluAlaGluGlyGlyAspLysAspAsnArgLeuLeuSerProArgProSerLeuLeuThrPr 947
DB 2483 CCAACCCGGAAGAGGAGGACAGGAGGTGCTGTCACCAAGGCCAGGCTCTCCACCCC 2542
QY 947 oThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuGlyLeuLeu 967
DB 2543 GGCTGAGATGCCCGGGCCAGTACCTCGCCCAAGAACCGCTGAGACTGAAGAGTGA 2602
QY 967 sGlnArgAlaAlaAlaLeuProProLleGlnValThrLysValHisGluProProArgG 987
DB 2603 GCAGCAGACGCGCGCATCCCTCCATC---GTCAACCAAGTTCATGAGCCCTCCGGGA 2659
QY 987 uAspAlaAlaProThrLysProAlaProProAlaProProProGlnAsnLeuGlnPr 1007
DB 2660 GGACACAGTACCCCAACCAAGTTCCTCCCTGCTCCACCCAGCAGACCTACAGCC 2719
QY 1007 oGluSerAspAlaProGlnGlnProGlySerSerProArgLysLysSerArgSerProAl 1027
DB 2720 AAGAGGTGACGTCTCAGCAGAGTCCGGAAGAGTCCAGGTGCAAGTCCCGCAGCCCACT 2779
QY 1027 aProProAlaAspLysGlu-----AlaPheAlaAlaGluAl 1039
DB 2780 GCCTCTGCGCGAAGAAAGAGCAGAAACCCGATCTTCCGCGCTTCCCACTAGG 2839
QY 1039 aGlnLysLeuProGlyAspProProCySerThrThrSerGlyLeuProPheProAlaProPr 1059
DB 2840 CCCAAAGCTACAGCTGAGGCCCGCCAGCGTGTATGAGGCTGCTCCCTCCCAATCCCTCC 2899
QY 1059 oArgGluValLleLysAlaSerProHisAlaProAspProSerAlaPheSerTyraLarPr 1079
DB 2900 ACGGAGGATGATTAACACTTCCCAACACCGCGCTGACCTTCCTCCACACACC 2959
QY 1079 oProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgPr 1099
DB 2960 CCCCCTGACACCGCTCTCTGGGCTCCACGATATGCGCGCTCCGCTCCACAGTCC 3019
QY 1099 oProThrLleSerAsnProProProLeuLleSerSerAlaLysHisProSerAlaLeuG 1119
DB 3020 CCCC---ATCTCTAACCCCCCACCCTCTCATCTCTGCGAACATGCCGCGCTACTGA 3076
QY 1119 uArgGlnLleGlyAlaLleSer---GlnGlyMetSerValGlnLeuHisValProTyse 1138
DB 3077 GAGCAGCTGGGCGCATCTCCACGACGAGGAGTGTCACTCACTTCGTGCTCACTC 3136
QY 1138 rGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspPr 1158
DB 3137 AGAGCATGCCAAGGCCCCCATGGGCTCTCAACATGGGGCTGCCCTTGCGTGAGACC 3196
QY 1158 oLysLysLeuAlaProPheSerLysValLysGlnGlnLeuLeuSerProArgLysGlnAl 1178
DB 3197 TAAAGAGCTG----- 3206
QY 1178 aGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgLysTh 1198
DB 3207 -----GGGAC 3211
QY 1198 rAlaLeuGlySerValProGlyGlySerLleThrLysGlyLleProSerThrArgValPr 1218
DB 3212 AGCAGTGGGCTCGGCACACGATGGAAGCATCAACCAAGGCGCTCCCAAGTACCGGGCTGC 3271
QY 1218 oSerAspSerAlaLleThrTyraArgLysSerLleThrHisGlyThrProAlaAspValle 1238
DB 3272 AGACGGCCCC-----AGCTACAGAGGCTCTATCACCCACGACGCGCCCAACGCTCT 3325
QY 1238 uTyraLysGlyThrLleThrArgLleLleGlyLysAspSerProSerArgLeuAspArgG 1258
DB 3326 CTCAAGAGGATCAATCAAGCATGCTGCTGAGACAGCCCAAGTGCCTTACCGGGC 3385
QY 1258 yArgGluAspSerLeuProLysGlyHisValLleTyraGluGlyLysLysGlyHisValle 1278

DB 3386 ACGAAGACACCCCGCCCAAGGCGCATGTCAATCATAGGGGCAAGAAAGCCACGCTCT 3445
QY 1278 uSerTyraGluGlyLysMetSerValThrGlnCySerSerLysGluAspGlyArgSerSe 1298
DB 3446 ATCTATGAAGGTGATGTCTGCTGTCAACAGTCTCTTAAGAGATGGAAGAGACGCTC 3505
QY 1298 rGlyProProHisGlyThrAlaAlaProLysArgThrTyraAspMetMetGlnGlyArgVa 1318
DB 3506 GGGCCCAACCCCATGAGACTGCGCCCTTAAGCACTATGATCATGATGAGAGGCGCTGT 3565
QY 1318 lGlyArgAlaLleSerSerAlaSerLleGlnLysMetGlyArgAlaLleProProG 1338
DB 3566 AGCAGAGACTGTACCTCAGCAGATAGAGGACTCATGGGCGCGCCATC---CTTA 3622
QY 1338 uArgHisSerProHisLleLysGlyGlnHisLleArgLysSerLleThrGlnG 1358
DB 3623 GCAGCACACCCC---CACCTCAAGAGAGAGATCATCAGAGGCTCATCAGCGAAG 3679
QY 1358 yLleProArgSerTyraGluAlaGlnGluAspTyraLeuArgArgGluAlaLysLeuLe 1378
DB 3680 CATCCCGAGTCTATGTGAGGCGCAGAGGACTTATACGCGGAGCGCAAGCTCTT 3739
QY 1378 uLysArgGluGlyThrProProProProProProSerArgAspLeuThrGluAlaTyTy 1398
DB 3740 GAAGCAGAGGAGACACACACCCACACACCTCGAGACTGTAGACCTACAA 3799
QY 1398 sThrGln-----AlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuVa 1415
DB 3800 GCGCCGCCCCCGACCCCTCTGGTCCCTGAAGCTGAAGCCGACCTACAGAGGTGTGT 3859
QY 1415 lAlaThrValLysGluAlaGlyArgSerLleHisGluLleProArgGluLeuArgHis 1435
DB 3860 ACGAATGTGAAGAGGAGCGCGGCTCTATCATGAGATCCGAGAGAGAGACTGGCGG 3919
QY 1435 sThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerLleThrGlnGlyTh 1455
DB 3920 CACACCTGAGTACCCCTGACACACAGGCTCTGAAGAGGATTCATATACCAAGGAC 3979
QY 1455 rProLeuLysTyraAspThrGlyAlaSerThrThrGlySerLysLysAspValArgSe 1475
DB 3980 CCACTCAAGTACGACTCGGGGACCCCTCAGTGGACCAAGAAACAGAGCTGGCTC 4039
QY 1475 rLeuLleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAlaAs 1495
DB 4040 CATATCGGCACACCCCGCGGCTTTCCCTCCCTGACCCGCTGAGATATGCTTA 4099
QY 1495 rAlaArgAlaLeuGluArgAlaCySerGluGluSerLeuLysSerArgProGlyThrAl 1515
DB 4100 CCGCCGAGCATGAGCGTGCCTGTATGAGAGAGTCTGAAGAGCCGATCAAGGACAG 4159
QY 1515 aSerSerSerGlyLysSerLleAlaArgLysAlaProValLleValProGluLeuGlyLys 1535
DB 4160 CAGTGTGACAGGGGCTCATCAACGTGGGCTCAGTCGTGCTCGAATGGGCAA 4219
QY 1535 aProArgGlnSerProLeuThrTyraGluAspHisGlyAlaProPheAlaGlyHisLeuPr 1555
DB 4220 GCCACGCGCAAAAGCCCACTGATTAAGAACACACGAGGACCCCTTCAACGATCACTTGC 4279
QY 1555 oArgGlySerProValThrMetArgLysProThrProArgLeuGlnGlySerLeuSe 1575
DB 4280 ACGTGGCTCCCTGTATCAACAGAGGAGCCACAGCCCTTCAAGAAAGCAGCTCT 4339
QY 1575 rSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluLleAlaLysSe 1595
DB 4340 ATCCAGCAAGCGCTCCCAAGACCGAAGCTGATCAACCCCGGAGATGCGCAAGTC 4399
QY 1595 rProHisSerThrValProGluHisHisProHisProLleSerProTyraGluHisLeuLe 1615
DB 4400 CCACACACACAGTGTCCGAGACACACCTCAACCTCATCTCCCTATGAGACACTTGTCT 4459
QY 1615 uArgGlyValLysGlyValAspLeuTyraArgSerHisLleProLeuAlaPheAspProTh 1635

QY	2350	MeHg1ylyerHytrApbG1nTPG1uGlUSeRProProleuSeR1AaAna1aPeAeNr	2370
Db	6617	CATGGGTAAATATGTCAGTGGGAAGACCCCCCGCGCTCGCGCCAAATGCTTTAAACC	6676
QY	2370	OLEuSnAlaSeR1ASeRLeuPro---AlaA1aMePPro1eThrAla1aASeP1yAR	2389
Db	6677	TCTGAATCCACGCCCACTCTGCCGCTGCTCTAATGCCATACCACTGCTGAACGACG	6736
QY	2389	qSeRAspH1eThrLeuThrSeRProG1yG1yG1yLysAla1yVa1SeRg1yA1gPr	2409
Db	6737	GAGTACACACCACTCACTCGCCAGGTGGAGGTGGAAAGCAAGTCTCTGGACGACC	6786
QY	2409	oSeRSeRrg1ySa1a1ySeRPro1A1ProG1yLeuAlaSeRg1yAaPa1gProProSe	2429
Db	6797	TAGACAGCGGAAAGCAAGTGCACGACCAAGCCCTACGCTCCGAGACCAACCCCTTC	6856
QY	2429	rVa1SeRSeRValH1SeRSeRg1uG1yAspCySaSnArgrThrPhro1euthrAsnArS1yVa	2449
Db	6857	TGCTCTCCCTCACTACACTAGAGGGGGAATTGCATATGCCAATCACCACCTACCAACCGTGT	6916
QY	2449	1TrpG1uAaPa1gProSeSeR1AglYSeRThrProBheProTyraaProleu1IeMe	2468
Db	6917	GTGGAGGAGACCGGCCCTCATCTGCAGGGGTCCACGCCATTCCTCCCTCAACCCCTTGATTAT	6976
QY	2469	tArgLeuG1nAlaG1yValMe1A1aSeRProProProG1yLeuPro1AglYSeRg1	2488
Db	6977	GAGGTACAGGACAGGTGTTCATGTGCTCCCGCCCACTGGCCTTGGGAGAGGACGGG	7036
QY	2489	yProleuAlaG1yProH1Sh1A1ATrAspG1uG1uPro1ySpProleuLeuCySeRg1	2509
Db	7037	GCCCTACTGCTGCTCCCAACCAAGCCTGGAGATGAGAGCCCAAGCCACTGCTGTGTTCACA	7086
QY	2509	nTyRg1uThrLeuSeRAspSeRg1u	2517
Db	7097	GTATGACACACTCTCGACAGCGAG	7121
RESULT 7			
LOCUS	CQ722208	6339 bp	DNA linear PAT 03-FEB-2004
DEFINITION	Sequence 8142 from Patent WO02068579.		
ACCESSION	CQ722208		
VERSION	CQ722208.1	GI:42283065	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS	1		
TITLE	Venter,C.J., Adams,M.C., Li,P.W. and Myers,B.W.		
	Kits, such as nucleic acid arrays, comprising a majority of		
	humanecons or transcripts, for detecting expression and other uses		
	thereof		
JOURNAL	Patent: WO 02068579-A, 8142 06-SEP-2002;		
FEATURES	pe Corporation (NY) (US)		
source	Location/Qualifiers		
	1..6339		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
ORIGIN			
Alignment Scores:			
Prod. No.:	2,39e-189	Length:	6339
Score:	9776.50	Matches:	1958
Percent Similarity:	75.36%	Conservative:	6
Best Local Similarity:	75.13%	Mismatch:	59
Query Match:	73.98%	Indels:	583
DB:	6	Gaps:	15
QY	US-09-522-753-5 (1-2517) x CQ722208 (1-6339)		
	1 MetSerGlySerThrGlnLeuValAlaGlnThrTPdAga1aThrGlnProArGTYrPro		20

Db	1	ATGTGGGATTCACACAGCGCTGTGGCACAAGACGTGAGGGGCACTGAGCCCCCGTACCG	60
QY	21	ProHisSerLeuSerTyrProValGlnIleAlaArgThriSthrAspValGlyLeuLeu	40
Db	61	CCCCACAGACCTTTCCTACCCAGTGCAGATGCGCCGGACGCACACGAGCGTGGGCTCTGT	120
QY	41	GluTyrGlnHisSsrAspArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln	60
Db	121	GAGTACCAAGCACCACTCCCGACTATAGCTTCCCACTGTGCGCCGCTTCATATCCAG	180
QY	61	ProGlnArgArgArgProSerLeuLeuSerGlnubheGlnProGlyAaGlnArgSerGln	80
Db	181	CCCCACGGCGCGAGAGCCCTCCCTCTGTCTGTAGTTCAGCGCCGGGAATGAACGTTCCAG	240
QY	81	GluLeuHisLeuArgProGlnSerHisSerTyrLeuProGluLeuGlySerGluMet	100
Db	241	GAGCTCCACCTCGCGGCAGAGTCCCACTCATATCCGCGCCGAGCTGGGGAGTCCAGATG	300
QY	101	GluubheIleGlnSerTyrArgProArgArgLeuGlnLeuLeuProAspProLeuLeuArgPro	120
Db	301	GAGTTTCATTTGAACCAAGCGCCCTCGCTAGAGTGTGCTGACCCCTGCTGCAGCG	360
QY	121	SerProLeuLeuAlaTargnGlyGlnProAlaGlySerGlnAspLeuThrLysAspArgSer	140
Db	361	TCAACCTCTCGTGGCCACGGGCGACCTCTGCGGAGCTGAAGCTCCACACAGAGACGTAGC	420
QY	141	LeuThrGlyLysLeuGlnProValSerProProSerProProHisSthrAspProGluLeu	160
Db	421	CTGACGGGCAAGCTGGAACCGGTCTCTCCCGCCCGCCGACACTGACCTCTGAGCTGT	480
QY	161	GluLeuValProProArgLeuSerTyrGlnGluLeuIleGlnAsnMetAspArgValAsp	180
Db	481	GAGCTGTGTGCGGCACAGCTGTCCAAAGAGAGCTGATCCAGAACATGACCGGTGGAC	540
QY	181	ArgGlnIleThrMetValGlnGlnGlnIleSerTyrLeuLysLysGlnGlnGlnLeu	200
Db	541	CGAAGATTCACCAATGATGAGACACAGATCTCTAGCTGAAGAAAGAGACAGCAACACTG	600
QY	201	GlnGlnGlnAlaAlaLysProProGlnProGlnLysProValSerProProIleGln	220
Db	601	GAGAGGAGGCTGCGCAAGCGCCCGAGCCTGAGAAGCCCGTGTACCGCGCCCATGTAG	660
QY	221	SerTyrHisSArgSerLeuValGlnIleIleTyrTrpAspGluMetAspGlyValGlnAla	240
Db	661	TCGAGACACCGCACCTCGTGTCAATATCATCAACAGAAACGGAAGAGCTGAAGCT	720
QY	241	AlaHisArgIleLeuGlnGlyLeuGlyProGlnAlaGluLeuProLeuTyrAsnGlnPro	260
Db	721	GCACATCGAATTCTGGAGAGGCTTGGGGCCCGAGTGAAGCTGCCCTGTACACCAAGCC	780
QY	261	SerAspThrArgGlnTyrHisSgluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu	280
Db	781	TCCGACACCGCGGAGTATCATGAGAACATCAAAATATAACGAGCGGATGCCGAAGAAGCTA	840
QY	281	IleLeuTyrPheLysValArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGln	300
Db	841	ATCTTGTAATCTCAAGAGAGGAATCAAGCTCGGAACAATGGGAGCAGAAAGTTCTGCAG	900
QY	301	ArgTyrTrpAspGlnLeuMetGlnAlaLeuGlnLysLysValGlnArgIleGlnAsnAsnPro	320
Db	901	CGCTATGACCAAGCTCATAGAGGCGCTGGGAGAGAAAGGTGAGCCCATCGAGACAAACCC	960
QY	321	ArgArgArgAlaLysGlnSerLysValArgGlnTyrTyrGlnLysGlnPheProGlnIle	340
Db	961	CGGCGCGCGCGCCAAAGAGAGAGAGTCCGCGAGTACTACGAGAAAGAGTTCCGAGATC	1020
QY	341	ArgLysGlnArgGluLeuGlnGlnArgMetGlnSerArgValGlyGlnArgGlySerGly	360
Db	1021	CGCAAGACAGCGCGAGCTGCAGAGAACCGCATGCAG--AGGTTGGGCGCAGCGGGCAGTGG	1077
QY	361	LeuSerMetSerAlaAlaArgSerGlnHisSgluValSerGlnIleIleAspGlyLeuSer	380
Db	1078	CTGTTCATGTGTGGCGCGCGGACGAGACGAGGTGTACAGATCATGTAGTGGCTCTCA	1137

QY 381 GUGINGIUAENLEUGIULYEGIMECARGIINLEUAIIVALLIIEPROMETLEUTYR 400
DB 1138 GAGCAGAGGAACCTGGAGAACAGATGGCGAGCTGGCCGTGATCCGCCCATCTGTGAC 1197
QY 401 ASPAIAASPGLINGINIAIGILEYSPHEIILEANMECHANGLYLEUMETALAASPPROMET 420
DB 1198 GACCTGACCGACGCGCATACAGTTTCATCAATGAACGGCTTATATGCGACCGACCATG 1257
QY 421 LYSVALTYRILYSAAPARGINVALMECASPMEETTPSERGIUGINGIULYSGIULTHPHE 440
DB 1258 AAGGTGACAAAGACCGCAGGTGATGAACATGTGGATGAGCGAGAGAAAGAACCTTC 1317
QY 441 ARGGLIULYSPHEMECTGINHIEPROLYSASPHEGLYLEUILEASERPHLEUGIULARG 460
DB 1318 CGGAGAGAGATTCAAGCAGCAGCATCCCAAGAACTTTGGCTGATCGCATCTCTCGAAGAG 1377
QY 461 LYETHRVALAGIUCYVALLEUTYRTYRILEUTHRILYSAENGLUAENTYRILYS 480
DB 1378 AAG---GTGAGTCGCTGC----- 1392
QY 481 SERLEUVALARGASERTYRARGARGGLIYSSERGINGLINGINGINGIN 500
DB 1392 ----- 1392
QY 501 GINGINGINGINGINGINGINGINPROMETPROARGSERSERGINGLIULYS 520
DB 1392 ----- 1392
QY 521 ASPGLIULYSGIULYSGIULVALAGIULYSGIULGLIULYSPROGLIULVALIGLUASN 540
DB 1393 -----CCGCCCATACCTCGCTCCACCC 1419
QY 541 ASPILYSGIULASPLEULEULYS-----GLULYETHRASPAPRTHRSERGIULYASASN 558
DB 1420 TCCGTGGGCAACTGCTGAGACTCAGGGAGAGAACACACACCTCAGGGAGAGAAC 1479
QY 559 ASPGLIULYSGIULVALAIAASERLYSGIYARGLYETHRALAASERGINGLIYARGARG 578
DB 1480 GACGAGAGAGAGGCTGTGGCTCCAAAGCCCGCAAACTCGCAACACCGAGAGAACGCC 1539
QY 579 LYSGLIYARGILETHRARGSERMETALAASNGIULAAASERGIULVALAILETHRPRO 598
DB 1540 AAAGGCGCGACACCCCTCAATGGCTAATGAGGCCAACAGCGAGAGCCATCACCC 1599
QY 599 GLINGISERIALAGIULASERMETALAASNGIULASERGIULSERERATGTPHGLIUL 618
DB 1600 CAGCAGAGCGCCGAGCTGGCTCCATGAGACTGAATGAGAGTTCTCGTGCACAGAGAA 1659
QY 619 GLUMECTULTHRILALYSGIYLEULEUGIULHIEGLIYARGAENTTPSERIALAIA 638
DB 1660 GAAATGAAACAGCCAGAAAGGTCTCTGAAACAGCGCCGCAACTGTGCGCATCGCC 1719
QY 639 ARGMEVALISERYSTRVALSERGINCYLYSASPHEPETHRPHASANTYRILYS 658
DB 1720 CGGATGTGGGCTCCAGACGTGTGCGAGGTGAAGAACTTCTCACTCAACAGAAAG 1779
QY 659 ARGGLIULASNEUMERGULILELEUGINGINHIELYLEULEULYSPHEGLIULYARGASN 678
DB 1780 AGGCAGAACTCATGATCTTGACAGACACAAAGCTGAAGATGGAGAGAGAGAAC 1839
QY 679 ALARGRGLYSLYSLYSLYSLYSLYSLYSLYSLYSLYSLYSLYSLYSLYSLYSLYSLY 698
DB 1840 GCCCGAG 1899
QY 699 VALVALIGIULASPGIULMETGLIULASERGIYVALSERGIYANGIULIULMETVAL 718
DB 1900 GTGTGTGAGATGAGATGAGATGAGAGCGTGGGCGGTGAGCGAAATGAGAGAGATGCTG 1959
QY 719 GLUGIULVALAGIULALEUHIISALASERGIYANGIULVALPROARGGLIULYSSERGIY 738
DB 1960 GAGAGGCTGAAGCTTAATGCTCTGGAGAAATGAG----- 1995

QY 739 PROALATHRVALASNANSENSESPRTHRGISERILEPROSERPROHISTHRIGLUALA 758
DB 1995 ----- 1995
QY 759 ALALYASAPRTHRGILYINANGIYPROLYSPROPROALATHRILEUGIYLAASPGIYPRO 778
DB 1995 ----- 1995
QY 779 PROPROGLYPROPROTHRPROPROARGRTHRSERARGALPROILEGLIUPROTHRPRO 798
DB 1995 ----- 1995
QY 799 ALASERGIULATHRGIYALAPROTHRPROPROPROALAPROPROSERPROSERALAPRO 818
DB 1995 ----- 1995
QY 819 PROPROVALVALPROLYSGIULYSGIULGLIULTHRILALALAPROPROVALIGLU 838
DB 1995 ----- 1995
QY 839 GLUGIYGLIULGLIULYSPROPROALALAGIULGLIULVALAASPTRHGLYLYSALA 858
DB 1995 ----- 1995
QY 859 GLUGIULPROVALIYSSERGIUCYETHRGLIULVALAGIULGLIULYPROALALYSGIYLYS 878
DB 1995 ----- 1995
QY 879 ASPALAGIULALAGIULATHRALAGIULYALALEULYSALAGIULYSGIULGLIY 898
DB 1995 ----- 1995
QY 899 GLYSERGIYARGALATHRTHRALALYSSERGIYALAPROGINASPSEASPSESER 918
DB 1996 -----GACACGACTCCAGT 2010
QY 919 ALATHRCYSESERIALAASPGIULVALAASPGIULAGIULGIYGLYASPLYSAENARGLEULEU 938
DB 2011 GCCACCTGCAGCTCAGACGAGGTGATGAGCGGAGGCGGCGCAAGAACCGGCTGCTG 2070
QY 939 SERPROARGPROSERILEULEUTHRPROTHRGIYASPPROARGALAAANALASERPROGIN 958
DB 2071 TCCCAAGGCGCCAGCTCTCCACCCGACGTGCGAGACCCCGGCGCAATGCTCACCCAG 2130
QY 959 LYSPEULEASPLEULEULYSGIULYSGIULNARGALALALALEPROCOLLEGINVAL 978
DB 2131 AAGCCACTGACCTGAAGCGCTGAAGCAGCGAGCGCTGCCATCCCCCATC----- 2184
QY 979 THRILYVALHIEGLIUPROPROARGIULASPALAIALPROTHRILYSPROALAPROPROALA 998
DB 2184 ----- 2184
QY 999 PROPROPROGINANLEUGINPROGLUSERASPALAPROGINPROGLYSSER 1018
DB 2184 ----- 2184
QY 1019 PROARGGLIYSESERARGSERPROALAPROPROALASPLYSGIULALPHEALALAGLU 1038
DB 2185 -----GAGGTGACGCTCTCGGCGCCCAAA----- 2211
QY 1039 ALAGIULYLEUPROGLYASPPROPCYSTRPHTHSERGIYLEUPROPHPROVALPRO 1058
DB 2212 -----TGACAC----- 2217
QY 1059 PROARGGLIULVALILEYLSALASERPROHIALAPROASPPROSERALAPHESERTYRALA 1078
DB 2218 -----GACATGTGC 2226
QY 1079 PROPROGLIYHIEPROLEUPROLEUGIYLEUHIISAPRTHRALAARGPROVALLEUPROARG 1098
DB 2227 TCCACAGGTCAACCATGCGCTCGGCGCTCCATGACAGTCCCGGCGGCTCGCGCGC 2286
QY 1099 PROPROTHRILASERANPROPROPROLEULIESERSERIALALYHIEPROSERVALLEU 1118

Db	2287	CCACCCACCATCTCAACCGCGCTCCCTCATCTCTGCAAGCAACCCAGGCTCTC	2346
Qy	1119	GIuArgInIleGlyAlaIleSerGInGlyMetSerValGInLeuHisValProTyrSer	1138
Db	2347	GAGAGGCAAAATAGGATGCATCTCCCA-----	2373
Qy	1139	GIuHisAlaValAlaProValIGlyProValThrMetGlyLeuProLeuProMetAspPro	1158
Db	2373	-----	2373
Qy	1159	LySylsLeuAlaProPheSerGlyValLySGInGInGInLeuSerProArgGlyGlnAla	1178
Db	2373	-----	2373
Qy	1179	GIyProProGInSerLeuGlyValProThrAlaGInGInAlaSerValLeuArgGlyThr	1198
Db	2373	-----	2373
Qy	1199	AlaLeuGlySerValProGlyGlySerIleThrLySGlyIleProSerThrArgValPro	1218
Db	2373	-----	2373
Qy	1219	SerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeu	1238
Db	2374	-----GGCAGCGCACGCTGACGCTCTG	2394
Qy	1239	TyrlLySGlyThrIleThrArgIleIleGlyGInAspSerProSerArgLeuAspArgGly	1258
Db	2395	ThCAAGGGACACATCACCGAGATCATCGCGCAGACACCCGAGTCGTTGGACCGCGC	2454
Qy	1259	ArgGInAspSerLeuProLySGlyHisValIleTyrGInGlyLyLyLySGlyHisValLeu	1278
Db	2455	CGGAGGACACGCTCGCCACAGGCGCACGTCATTCAGAAAGGCAAGAGGCGCACGTCCTG	2514
Qy	1279	SerTyrGInGlyGlyMetSerValIleThrGInCySerSerLySGInAspGlyArgSerSerSer	1298
Db	2515	TCCATATAGAGGTGCGATCTCTGTGACCCAGTGCTCCAAAGAGGACGCGCAAGACGCTCA	2574
Qy	1299	GIyProProHisGlyThrAlaAlaProLyAspArgThrTyrAspMetMetGInGlyArgVal	1318
Db	2575	GGACCCCCCATGAGACGCGCGCCCCCAAGCGCACTTATGACATATGAGAGGCGCGGTG	2634
Qy	1319	GIyArgAlaIleSerSerAlaSerIleGInGlyLeuMetGlyArgAlaIleProProGIn	1338
Db	2635	GGCAGAGGCATCTCTCTCAGCAGCATGCAAGGT-----	2667
Qy	1339	ArgHisSerProHisLeuLeuGInGInHisHisIleArgGlySerIleThrGInGly	1358
Db	2667	-----	2667
Qy	1359	IleProArgSerTyrValGInAlaGInGInAspTyrLeuArgArgGInAlaLyLeuLeu	1378
Db	2668	-----GATACAGAGGAGAGACATTCACTCTC	2694
Qy	1379	LySArgGInGlyThrProProProProProProSerArgAspLeuThrGInAlaTyrLyS	1398
Db	2695	-----GGTGGCCCC-----TGGTGGCGGCGCTACAG	2721
Qy	1399	ThrGInAlaLeuGlyProLeuLyLeuLyAspProAlaHisGlyGlyLeuValAlaIleThrVal	1418
Db	2722	ACGCAAGGCGCTCGGCGCCCTGAAGCTGAAGCGCGCCATGAGGGCGCTGTGTGCCACGGTG	2781
Qy	1419	LySGInAlaGlyArgSerIleHisGInIleProArgGInGInLyLeuArgHisIleThrProGIn	1438
Db	2782	AAGAGAGGGGCGCGCTCCATCATGATGATCCCGCGGAGGCTGGCGGACACGCCCGAG	2841
Qy	1439	LeuProLeuAlaProArgProProLeuLySGInGlySerIleThrGInGlyThrProLeuLyS	1458
Db	2842	CTGCGCCCTCGGCGCGCGCGCTCAAGAGGGCTTCATACGCAAGGACACCCCGCTCAAG	2901
Qy	1459	TyrAspThrGlyAlaSerThrThrGlySerLyLyHisAspValArgSerLeuIleGly	1478

Db	2902	TACGACACCGGCGGCTCCACCACTGGCTCCAAAAGACAGACGTAAGCTCCCTCATGGCG	2961
Qy	1479	SeSPProGlyaThrPheProProValHisProLeuAspValMetAlaAspAlaArgAla	1498
Db	2962	AGCCCCGGCCGGAGGTTCCACCCTGGTCACCCCTGGATGTGAAGGGCCGACCGCCGGGACA	3021
Qy	1449	LeuGluArgAlaCysGlyArgGluSerLeuLysSerArgProGlyThrAlaSerSerSer	1518
Db	3022	CTGGAACCTGCTCTACAGAGAAAGCTTGAAGAGCCGGCCAGGAGACGGCCAGCTCG	3081
Qy	1519	GlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgGln	1538
Db	3082	GGGGGCTCCATTGGCGCGGGCGCCCGGTGCTATTGGCTGAGCTGGGCAAGCGCGGAG	3141
Qy	1539	SeSPProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySer	1558
Db	3142	AGCCCCCTTAACCTTGAAGAGCAAGGGGACCCCTTTCGGGCAACCTCCACGAGGTTCG	3201
Qy	1559	ProValThrMetAlaArgGluProThrProArgLeuGlnGlySerLeuSerSerSerLys	1578
Db	3202	CCCCGTACACCGGGAGCCCAACCCCGCGCTGGAGAGGGCAACCTTTCGTCAAGCAAG	3261
Qy	1579	AlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSer	1598
Db	3262	GCAATCCCAAGACCCAAACTGACTGACGCTGTGATGATGCCAATGCCCCACACAGC	3321
Qy	1599	ThrValProGluHisAspProHisProIleSerProTyrGluHisLeuLeuArgGlyVal	1618
Db	3322	ACCGTGCCGAGACACACCCCAACCCCATCTGCGCTATGAGCACCTGCTCGGGCGTG	3381
Qy	1619	SerGlyValAspLeuTyrArgSerHisGlyProLeuAlaPheAspProThrSerIlePro	1638
Db	3382	AGTGCCGGGACCTGTATCCGAGCAATATCCCTGGCGCTTGCACCCCACTCATATCC	3441
Qy	1639	ArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsn	1658
Db	3442	CGCGGCACTCCCTTGAGAGCAGCCGCTGCTACTACCTGCCCCGACACTGGCCCCAAC	3501
Qy	1659	ProThrTyrProHisLeuTyrProProTyrIleuIleArgGlyTyrProAspThrAlaAla	1678
Db	3502	CCCAACTACCGGCACTGTATCCACCCCTACTATCCGGGCTACCCGCAACGGCGGCG	3561
Qy	1679	LeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisAsn	1698
Db	3562	CTGGAGAACCGCGAGACCATCATCAATGACTACATCACTCGACGAGATGCACACAAAC	3621
Qy	1699	ThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgLysSerProArgGlySer	1718
Db	3622	GCGGCAACCGGCATAGGCCCAACGAGCTGATGTCTGAGGGGCTCTGCCCCGGAGTCC	3681
Qy	1719	SerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValPro	1738
Db	3682	TGCGTGGCACTCAACTACGTGCGGGGTCCCGAAGCATCACTGATCCCAAGTGCA	3741
Qy	1739	HisLeuProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeu	1758
Db	3742	CACCTGCTGTGCTGTGCCCCCGACACACAGGACCCGACGCAACGCGCATGACGCGCTT	3801
Qy	1759	AlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSer	1778
Db	3802	GCTTACCTCCCGACCGCGCCCGACCCCTTCAGACCGCCCAACAGATCCCACTCC	3861
Qy	1779	ProGlyGlyProThrHisLeuThrLysAspProThrThrThrSerSerSerGluValArg	1798
Db	3862	CCAGGAGTCCAAACACTTACAAAAACCAACACACAGTCTCTGTCGAGCGGAGGAGA	3921
Qy	1799	AspArgAspArgGluArgAspArgAspArgGluArgLysSerIleLeuThrSerThr	1818
Db	3922	GACCGGAGTCAAGAGCGGAGCCGGGATCGGAGCGGAAAGTCAATCTCTACATCCAC	3981
Qy	1819	ThrThrValGluHisAlaProIleTyrArgProGlyThrGluGlnSerSerGlySerSer	1838
Db	3982	ACGACGGTGAAGACACGCCCATCTGGAGACTGTGTAGGCGCATGAGACCCCAACCCCC	4041

QY 1839 GlySerSerGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHis 1858
 DB 4442 GCTCCGGAGACTCTTGTGGCCGCAAGAGGCTCCCTGCTGTAATGCCACTGACTGTAC 4101
 QY 1859 GlnHisSerProIleSerProArgThrGlnAapAlaLeuGlnGlnArgProSerValLeu 1878
 DB 4102 CAGACTCGCCACTCTCCCTCGGACCCAGATGCCCTCCAGCAGAGACCAAGTGTCTT 4161
 QY 1879 HisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeu 1898
 DB 4162 CACAAACAGGCAAGAGGATATCATCACCGCTGGAGGCCAGCAGCCAGGCTCTG 4221
 QY 1899 ArgSerThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHis 1918
 DB 4222 AGGTCCACTCTCACTCTCACTCCGTTGCGCCGCTGCACATTCACACTGCCACACCAC 4281
 QY 1919 CysArgProLeuGlyGlyThrLeuAapGlyValTyrrProThrLeuMetGluProValLeuLeu 1938
 DB 4282 TGCCCACTGGGCGGCGACCTCGATGGGGTCTACCTTCACTGAGAGCCGCTCTGCTG 4341
 QY 1939 ProLysGlnAlaProArgValAlaArgProGluArgProAlaArgAlaAapThrGlyHisAla 1958
 DB 4342 CCCAAGAGGCCCCCGGGTGCCTGGCCAGAGCGGCCCGGAGACACACCGGCAATGCC 4401
 QY 1959 PheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGly 1978
 DB 4402 TTCTTCCCAAGACCCCAAGCCGCTCGGGCTGAGAGCCGCTCTCTCCCAAGCAAGGCG 4461
 QY 1979 SerGluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrPro 1998
 DB 4462 TCGGAGCGCGGCGCTTATGTCCTCTGTCTGTGGCCACGCGCACATCGCCCGCACCTT 4521
 QY 1999 AlaLysAsnLeuAlaProHisHisAlaSerProAapProProAlaProProAlaSerAla 2018
 DB 4522 GCGAAGAACCTCGCACTCAACAGCAGCGCCGAGCCGCGCGCCACCTGCTCGGCC 4581
 QY 2019 SerAapProHisArgGlyLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeu 2038
 DB 4582 TCGGACCCGCGACCGGAAAGACTCAAGTAACCTTTTCCATCCAGAACTGGAATCTC 4641
 QY 2039 ArgSerLeu----- 2041
 DB 4642 CGTTCTCTGGGTAAAGCACCCCTGACAGCGGCACCTTCATAGACGATATCATGCTGT 4701
 QY 2041 ----- 2041
 DB 4702 CAAATTGCTCAGATAAAGGGCGCGAAGAGAGTGCCTGGCCAAAGGCTCCCTCGC 4761
 QY 2042 ---GlyTyrrHisGlySerSerTyrrSerProGluGlyValGluProValSerProValSer 2060
 DB 4762 GATGGTTATCAACGCGACAGCTACAGCCCGGAGGGGTGAGAGCCGTCAGCTGTAGAC 4821
 QY 2061 SerProSerLeuThrHisAapLysGlyLeuProLysHisLeuGluGluLeuAapLysSer 2080
 DB 4822 TCACCCAGCTGACCCAGCAAGAGGGCTCCCAAGCACTCGAAGAGCTCGACAAAGC 4881
 QY 2081 HisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGlyGluAla 2100
 DB 4882 CACTGGAGGGGAGCTGCGGCCCAAGCAGCAGCCCGGTGAAGCTTGGCGGGAGGCC 4941
 QY 2101 AlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSerProLeuLeu 2120
 DB 4942 GCCCACTTCCACACTCGCGGCGCTCTGAGAGCCAGCCCTGTCTCAAGCCGCTCTCTC 5001
 QY 2121 GlnThrAlaProGlyValLysGlyHisGlnArgValAlaThrLeuAlaGlnHisIleSer 2140
 DB 5002 CAGAGCCGCCCAAGGATCAAAAGTCAACAGCGGGTGTCACTGTGCCAGCATCATAGT 5061
 QY 2140 ----- 2140
 DB 5062 GTAACTACGGGTTCTGTGCTGTGCTTGTCACTTTTGACCTGGGGGACCAAGCCGTGGA 5121

QY 2140 ----- 2140
 DB 5122 GAGGGATGGGGAACCCACAGGCCCTTCTGTCTGTGGGGGTGGCTGGGGATTCAGGGC 5181
 QY 2141 -----GluValIleThrGlnAapTyrrThrArgHisHis 2151
 DB 5182 ATGGCGCTGGGGGATTCAGAGGGCGTGGAGGTATCAACAGGACTTACACCCGACACAC 5241
 QY 2152 ProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrrSerPheProGlyValaSerCys 2171
 DB 5242 CCACAGAGCTCAGCGACCCCTGCGCCGCCCTCTTACTCTTCTCTGGGGCCAGCTGC 5301
 QY 2172 ProValLeuAapLeuArgArgProProSerAapLeuTyrrLeuProProProAapHisGly 2191
 DB 5302 CCGTCTTGAGACTCCGCGGCCCAACAGAGACTTCACTTCCGCGCCCGGACATAGT 5361
 QY 2192 AlaProAlaArgGlySerProHisSerGlyGlyLysArgSerProGluProAapLys 2211
 DB 5362 GCCCGGCGCGTGTCTCCCCCAAGAGAGGGGGCAAGAGTCTCCAGAGCCAAACAG 5421
 QY 2212 ThrSerValLeuGlyGlyGlyLysArgGlyIleGluProValSerProProGluGlyMet 2231
 DB 5422 ACGTGGTCTTGGTGTGTGTGAGAGCGATTAATGACTGTGTCCACCGAGGGCATG 5481
 QY 2232 ThrGluProGlyHisSerArgSerAlaValTyrrProLeuLeuTyrrArgAapGlyGluGln 2251
 DB 5482 ACGGAGCGACGACACTCCCGAGTGTGTGTACCGCTGTATCCGGATGGGGAAACAG 5541
 QY 2252 ThrGluProSerArgMetGlySerLysSerProGlyAapThrSerGlnProProAlaPhe 2271
 DB 5542 ACGGAGGCC---AGATGGGGCTCCAAGTCTCAGGCAACACAGCAGCGCCAGCCTTC 5598
 QY 2272 PheSerLysLeuThrGlySerAapSerAlaMetValLysSerLysLysGlnGluIleAap 2291
 DB 5599 TTCAGCAAGCTGACCGAGAACCTCCGCAATGCTCAAGTCAAGAGCAAGAGATCAAC 5658
 QY 2292 LysLeuLeuAapThrHisAapArgAapGluProGlyTyrrAapIleSerGlnProGlyThr 2311
 DB 5659 AAGAGCTGAACACCAACACCGGAATGAGCTCGAATACATATCAACCGACTGGGACG 5718
 QY 2312 GluIlePheAapMetProAlaIleThrGlyThrGlyLeuMetThrTyrrArgSerGlnAla 2331
 DB 5719 GAGATCTTCAATATGCGCCGCTATCACCAGAACGCGCTTATGACTATTAAGACCAAGCGC 5778
 QY 2332 ValGlnGlnHisAlaSerThrAapMetGlyLeuGluAlaIleIleArgLysAlaLeuMet 2351
 DB 5779 GTCCAGAGACATCCAGACCAACATGAGGCTGGAAGCCATTAATTAAGAGGCACTCATG 5838
 QY 2352 GlyLysTyrrAapGlnTyrrGluGluSerProProLeuSerAlaAapAlaPheAapProLeu 2371
 DB 5839 GGTAAATATGACAGTGGGAAGAGTCCCGCGGCTCAGCGCCAAATGCTTTAACCTCTG 5898
 QY 2372 AapAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAapGlyArgSerAap 2391
 DB 5899 AATGCCAGTGCACGCTCGCCGCTGTATGCCATTAACGCTGTGACGAGCAGAGTAC 5958
 QY 2392 HisThrLeuThrSerProGlyGlyGlyLysValLysValSerGlyArgProSerSer 2411
 DB 5959 CACACACTCACTCGCCAGGTGGCGGGAGAGGCAAGGTCTCTGCAAGACCCAGCAGC 6018
 QY 2412 ArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAapArgProProSerValSer 2431
 DB 6019 CGAAAGCCCAAGTCCCGCGGCCGCTGGCATCTGGGAGACGGGCACTCTGTCTCC 6078
 QY 2432 SerValHisSerGlyLysProCysAapArgArgThrProLeuThrAapArgValTyrrGlu 2451
 DB 6079 TCAGTCACTCGGAGGAGACATGCAACCGCGGACCCCTCAACCAACCGCGTGGGAG 6138
 QY 2452 AapArgProSerSerArgLysSerThrProPheProTyrrAapProLeuIleMetArgLeu 2471
 DB 6139 GACAGGCCCTGTCTGCCAGGTTCACGCCATTCCTTCAACACCCCTGTATCAAGCGGCTG 6198
 QY 2472 GlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeu 2491

[illegible]

exon
 4483..4620
 /gene="SMRT"
 /note="alternatively spliced insert"
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,33e-151 Length: 5989
 Score: 7885.00 Matches: 1510
 Percent Similarity: 94.34% Conservative: 7
 Best Local Similarity: 93.91% Mismatches: 14
 Query Match: 59.67% Indels: 77
 DB: 9 Gaps: 7
 US-09-522-753-5 (1-2517) x HSU37146 (1-5989)
 QY 984 ProProArgGluAspAlaAlaProThrLysProAlaProProAla----- 998
 |||||
 DB 168 CCACCAAGA-----TCACCATATAGCCACGACCCCTCGATAGATAATCATAT 218
 999 ProProProPro----- 1002
 |||||
 DB 219 CCGAGCCACCCCTGTGTTCTAATGACCATCTGCCATTCTATGGCTCCATGTATCTATT 278
 1003 -----GlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGly 1016
 |||||
 DB 279 GCAGAGTCCCAAGTTTTCATCAGCACTGMAACCAACATGCCGTGTAAAGGTCTCCAGT 318
 1017 -----SerSerPro-----Arg 1020D
 |||||
 DB 339 CTCCCTCTTCTGTATAGGCCCATGAGGTCTCCCTCTAAGTGTGTGTGTGAGAGAAGA 398
 1021 GlyLysSerArgSer-----ProAlaProPro-Ala----- 1030D
 |||||
 DB 399 GGAAGATTTCGAGCTCTGCTACCAACCCCGGCGCTGTGTCCGGAGAGAGGGAGGTGT 458
 1030 ----- 1030D
 459 CCTGACGAGGGCTCATATCATCTAATCTTCTAATGAGGACATGGAGCGGCCACCC 518
 1031 -AspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCysSTPTrh 1050D
 519 CGACAAAGAGGCTTCGCGAGCGCGAGCCCGAAGAACTGCTGGGGACCCCTTCTCTGGAC 578
 1050 rSerGlyLeuProPheProValProProArgGluValIleLysAlaSerProHisAlaPr 1070D
 579 TTCGGCGCTTGGCTTCCCGCGGCCCGCGAGAGTATCAAGGCTTCCCGCATGGCCCC 638
 1070 oASPProSerAlaPheSerTyrAlaProProGlyHisAProLeuProLeuGlyLeuHisAs 1090D
 639 GGACCCCTCAACCTTCTCTAAGCTCCACCTGTGCACCACTGCCCTTGGGCTTCATGA 698
 1090 pThrAlaArgProValLeuProArgProProThrIleSerAsnProProLeuIleSe 1110D
 699 CACTGCCCGCGCGCTCGCGCGGCCGCCACCATCTCCAACCGGCTCCCTCATCTC 758
 1110 rSerAlaLysHisProSerValLeuGluValGlnIleGlyValIleSerGlnGlyMetSe 1130D
 759 CTGTGCCAAGACACCCACGCTCTTCGAGAGGCAATAGTGCCATCTCCCAAGAGAAATGC 818
 1130 rValGlnLeuHisValProTyrSerGlnHisAlaIleValAlaProValGlyProValThrMe 1150D
 819 GATCCAGCTCCACGTCCTCGTACTAAGACATGCCAAGCCCCGGTGGGCTCTGCACCAT 878
 1150 tGlyLeuProLeuProMetAspProTyrLysLeuAlaProPheSerGlyValLysGlnG 1170D
 879 GGGGCTGGCCCTGGCCCATGAGACCCCAAAAGCTGGACCTTTGACGGAGTGAAGCAGGA 938
 1170 uGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaG 1190D

Db 939 GCAGCTTCCCGACGGGGCCAGGCTGGGGCCACCGGAGAGCTGGGGGTGGCCACAGGCCA 998
Qy 1190 ngluaSerValIleuArgIyThraIaleuGlySerValProGIyGlySerIleThrIy 1210
Db 999 GGAGGCTCCGTGCTTGAGAGGAGACAGCTTGAGGCTCAAGTTCCGGGGGAGACATACCAA 1058
Qy 1210 sglIyIleProSerThraArgValProSerAaspSerAlIleThrIyArGlySerIleTh 1230
Db 1059 AGGCATTCACAGCACCGGGGGCCCTCGACAGGCCATCAATACCGGGCTCATCAC 1118
Qy 1230 rHIsGlyThProAlaaspValIleuTyLyGlyThIleThraArgIleIleGIyIuas 1250
Db 1119 CCACGGGACCGCAGCTACGCTCTGTACAAGGACACCATCACCGAGTCAATCCGCGAGGA 1178
Qy 1250 pSerProSerArGleuAaspArgIyArGlyIuaSerIleuProIyagIyIleValIleTy 1270
Db 1179 CAGCCGAGTCGCTTGAGACGGGGCCGGAGAGACAGCTCGCCCAAGGGCCACGTCATCTA 1238
Qy 1270 rGIuGIyLyIeYsGIyHIsValIleuSerTyrGIuGIyGIyMeSerValThrGInCySe 1290
Db 1239 CGAAGGCAABAAGGCCACGCTTGTCTTAAGAGGTGGCATGTCTTGACCAAGTCCTC 1298
Qy 1290 rLyGlyAaspGIyArGySerSerGIyProProHIsGIuThraIaIaProIyAargTh 1310
Db 1299 CAAAGAGAGACGAGAAAGAGCTCAGAGACCCCCATGAGACGGCCGCCCAAGCGCAC 1358
Qy 1310 rTyraaPmeMeGIuGIyArGlyArGlyAlIleSerSerAlaSerIleGIuGIyIe 1330
Db 1359 CTATGACATGATGAGAGGGCCGGGGGAGAGAGCCATCTCTCAAGCCAGCATCAAGGTCT 1418
Qy 1330 uMeGIyArGlyAlIleProProGIuArGHIIsSerProHIsHIsleuYsGIuGInHIsH 1350
Db 1419 CATGGGCGGTGCATCCCGCGAGAGGACACAGCCCCACCATCTCAAAAGACAGACCA 1478
Qy 1350 sIIleArGlySerIleThrGInGIyIleProArGySerTyrValGIuAlaGInGIuAapTy 1370
Db 1479 CATCCCGGGTTCATACACAAAGAGATCCCTCGTCTTAAGAGGACAGAGAGACTA 1538
Qy 1370 rIleuArGyArGlyAlaIleuAyleuIeIuYsArGyGIuGIyThProProProProPse 1390
Db 1539 CCGGCTCGGAGGCCAAAGCTCTTAAGCGGAGGACCGCTCCGCCCCACCGGCTTC 1598
Qy 1390 rArGaaPleuThrGIuAlaTyLyThrGInAlaIeGIyProleuYsleuYsProAl 1410
Db 1599 ACGGACCTGAACCGAGGCTTAAGACGAGGCCCTGGGGCCCTGAAGCTGAAGCCGGC 1658
Qy 1410 sHIsGIuGIyLeuValAlaThrValIyGlyIuaGIyArGySerIleHIsGIuIleProAr 1430
Db 1659 CCATGAGGGCTGTGGCCACGGTGAAAGAGCGGGCGGCTCATGCATGATCCCGCG 1718
Qy 1430 gGIuGIuIeAargHIsThProGIuIeupProleuAlaProArGyProleuYsGIuGIySe 1450
Db 1719 CGAGGAGCTGGGACACAGCCGAGCTGCCCCCTGGGCCCGGCTCAAGAGGGGCTC 1778
Qy 1450 rIleThrGInGIyThProleuYsTyraaPThrGIyAlaSerThThrGIySerIySly 1470
Db 1779 CATCAAGCAGAGGACCCCGCTCAAGTACACACCGGCGCTCAACATGGCTCCAAA 1838
Qy 1470 sHIsaapValArgSerIleuIleGIySerProGIyArGThrPheProProValHIsaProIe 1490
Db 1839 GCACGACGTACGCTCCCTCATCGGAGCCCGGCGGACGTTCCCAACCGGTGACCCGCT 1898
Qy 1490 uAaPValMeAlaAaPAlaArgAlaIeGIuIaArgAlaCyTyrGIuGIuSerIeIySse 1510
Db 1899 GGAATGTAATGGCCGAGCGCCGGGCACTGGAACGTGCTGTACAGAGAGAGCTGAAGAG 1958
Qy 1510 rArGProGIyThraIaSerSerSerGIySerIleAlaArgGIyAlaProValIleVa 1530
Db 1959 CCGGCAAGGAGCCGACAGACGCTCGGGGGGCTCATTTGGCGGGGGCCCGGCTCATTTGT 2018
Qy 1530 IProGIuIeGIyLyAaPArGInSerProleuThrTyrGIuAaPHIIsGIyAlaProPh 1550

Db 2019 GCCTGAGCTGGTAAGCCGGGACAGAGCCCTGACTATAGGACCACGGGACCCCTT 2078
Qy 1550 eAlaGIyHIsIeupProArGlySerProValThMeArGIuIProthProArGIuGI 1570
Db 2079 TGCCGGCCACTCCACAGAGATTCGCCGTGACCATAGCGGAGGCCACGGCGGCTGCA 2138
Qy 1570 nglIySerIeIuSerSerSerIySAlaSerGIaAaPArgLyIleuThSerThProAr 1590
Db 2139 GAGAGGAGCCTTTGCTTCAGCAAGGATCCAGAACCGAAAGGTGACGTGACGCTCG 2198
Qy 1590 gGIuIleAlaIySerProHIsSerThraIProGIuHIsHIsaProIleSerPr 1610
Db 2199 TGAGATCGCAAGTCCCGGACAGACCCGCGGAGCACACCCACCTTCGCGC 2258
Qy 1610 oTyrGIuHIsIeIuAaArgIyValaSerGIyValaAaPleuTyraaSerHIsIleProIe 1630
Db 2259 CTATGACACTGCTTCGGGGCGTGAATGGCGTGAACCTGTATCGGACCAATCCCCCT 2318
Qy 1630 uAlaPaaPProThSerIleProArGIyIleProleuAaPAlaAlaIaIaTyTy 1650
Db 2319 GGCCTTGACCCCACTCCATACCCCGGCGCATCCCTTGAGAGCAGCGGCTGCTACTA 2378
Qy 1650 rIeupProArGHIIsIeuaIaProAaPProThTyProHIsIleuTyProProTyIeui 1670
Db 2379 CTTGCCCGGACACTGGGCCCAACCCCACTACCGGACCTGTATCCACCTTACTAT 2438
Qy 1670 eArGyIyTyProaPThraIaIeGIuIaAaArgInThrIleIleAaPArTyI 1690
Db 2439 CCGGCTTACCCGACAGCGGGCGGTGAGAAACGGAGAACCATATACATATACATAT 2498
Qy 1690 eThrSerGIuIeMeCHIsHIsaThraIaThraIaMeAlaGIaArgAlaAaPMeCle 1710
Db 2499 CACCTGACAGAGATGACACACAAAGGACACCGGCGCATGGCCAGAGCTATATGCT 2558
Qy 1710 uArGIyIeIuSerProArGIuSerSerIeuaIaIeuaPThraIaIaGIyProArGI 1730
Db 2559 GAGGGGCTCTCGCCCGGAGAGTCTCGGTGACATCACTAGCGTGGGGTCCGAGG 2618
Qy 1730 yIIleIleAaPleuSerGIuValProHIsIeupProValIleuValProProthProGIyTh 1750
Db 2619 CATCATGACACTGTCCCAAGTGCACACCTGCTGTGTCGTCGCCCGACACAGCAC 2678
Qy 1750 rProAlaThraIaMeArAaPArgIeuaIaTyIeupProThraIaProGIaProPheSerSe 1770
Db 2679 CCGAGCCACCGCATGAGACGCTTGTCTTACCTCCCAACCGGCCCCACGCTTACGAG 2738
Qy 1770 rArGHIIsSerSerSerProleuSerProGIyIyProthHIsIleuThIyAaPProThTh 1790
Db 2739 CCGCCACAGACGCTCCCACTCTCCCAAGAGGTCCAAACACTTGACAAACCAACAC 2798
Qy 1790 rThrSerSerSerGIuArGIuIaArgAaPArgAaPArgIyIuaIaArgAaPArgGIuAr 1810
Db 2799 CAGTCTCTGTCCGAGCGGAGACGAGACCGGATTCAGAGCGGAGACCGGATTCGAGAG 2858
Qy 1810 gGIuYsSerIleIeThSerThThThraValGIuHIsAlaProIleTyraaPProGI 1830
Db 2859 GGAaAAGTCACTCTCACTTCAACAGACGAGTGAAGACGACCACTTGAAGACTGG 2918
Qy 1830 yThrGIuIuSerSerGIySerSerGIySerSerGIyGIyGIyGIySerSerAr 1850
Db 2919 TACAGAGCAGACAGCGGACGAGCGGAGCGGCGGGGTGGGGGAGCAGCAGCGCG 2978
Qy 1850 gProAlaSerHIsSerHIsAlaHIsGIuHIsaPleuIleSerProArGThGIaAaPAl 1870
Db 2979 CCGGCTTCCACTCCATGACCCACAGACATCGCCCATCTCCCTCGGACCCAGGATCC 3038
Qy 1870 aIeGIuGIaArgProSerValIleuHIsaThraIaMeCleYsGIyIleIleThraIaVa 1890
Db 3039 CCTCAGAGAGAACCAAGTGTCTTCAACAACAGGCAAGAGGTATATACACGCTGT 3098
Qy 1890 IGIuProSerTyProThValIleuAaPArgSerThSerThSerSerProValArgProAl 1910
Db 3099 GGAGCCACAGAACCCACGCTCTGAGGTCCACTTCCACTTCCACTTCCGCTTCCGCGCAC 3158

QY 1910 aAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValIleTyrPr 1930
 Db 3159 TGCCACATTCCTCCACCTGCCACCACTGCCCACTGGCGGCACTCTCGATGGGGTCTAACCC 3218
 QY 1930 oThrLeuMetGluProValLeuLeuProGlyGlyAlaProArgValAlaArgProGluAr 1950
 Db 3219 TACCTCTCAATGAGACCGCTCTGCTGCTCCCAAGAGAGGCCCCGGGCTGCCGCGCAGAGAG 3278
 QY 1950 gProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGly 1970
 Db 3279 GCGCCGAGAGACACCGCGCATGCTCTCTCGCAAGCCCCCAAGCCGCTCGGGCTGGA 3338
 QY 1970 uProAlaSerSerProSerLeuGlySerGluProArgProLeuValProProValSerGly 1990
 Db 3339 GCGCCGCTCTCTCCCGCAGAAAGGCTCGAAGCCCCGCGCTAGTGCTCTCTGCTCTG 3398
 QY 1990 yHisAlaThrIleAlaArgThrProAlaLysAlaLeuAlaProHisAlaSerProAs 2010
 Db 3399 CCACGCGCACTGCGCGCACCCCTCGAAGAACTCGACCTCACCGCGCAGCCGGA 3458
 QY 2010 pProProAlaProProAlaSerAlaSerProHisArgGlyLysThrGlnSerLysPr 2030
 Db 3459 CCGCGCGGCGCACTGCTCGGCTCGGACCGGACCGGAAAGACTCAAGTAAGTAACC 3518
 QY 2030 oPheSerIleGlnGluLeuGlyLeuArgSerLeuGlyTyrHisGlySerSerTyrSerPr 2050
 Db 3519 CTTTTCATCCAGAACTGGAACTCCGTTCTCGGGTTACACGCGAGCAGCTACAGCC 3578
 QY 2050 oGluGlyValGluProValSerProValSerProSerLeuThrHisAspLysGlyLe 2070
 Db 3579 CGAAGGGGTGAGCCCGTGAAGCCCTGTGAGCTCACCAAGTCTGACCAAGAGGGGCT 3638
 QY 2070 uProLysHisLeuGluLeuAspLysSerHisLeuGlyGlyValLeuArgProLysGly 2090
 Db 3639 CCCCAAGCACTTGGAAGACTCGACAAAGCACTGAGAGGGGAGCTGGGCGCAAGGA 3698
 QY 2090 nProGlyProValLysLeuGlyGlyValAlaHisLeuProHisLeuArgProLeuPr 2110
 Db 3699 GCCAGGCCCCGTGAAGCTTGCGGGAGGCGGCCCACTCCCACTGCGGCGCTGCGC 3758
 QY 2110 oGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGly 2130
 Db 3759 TGAGAGCCAGCCCTCGTCAGCCGCTGCTCCAGACCGGCCCAAGGGGTCAAGAGTCAACA 3818
 QY 2130 nArgValValThrLeuAlaGlnHisIleSerGlyValIleThrGlnAspThrThrArgHis 2150
 Db 3819 GCGGGGTGTACCTTGAGCCCAAGCAATCAGTGAAGTCAACACAGAACTACACCGGGA 3878
 QY 2150 sHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSe 2170
 Db 3879 CCAACCAAGCACTCAGGCGCACTCTGCCGCGCCCTCTAATCTCTTCCCTGGGCGCAG 3938
 QY 2170 rCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHis 2190
 Db 3939 CTGCCCGCTGCTGAGACTCGCGCGCCCACTCACTGACTCTCCGCGCCGAGCA 3998
 QY 2190 sGlyAlaProAlaArgGlySerProHisSerGlyGlyGlyLysArgSerProGluProAs 2210
 Db 3999 TGCTGCCCGCGCGCGCTGCTCCCCCAAGCAAGGGGGGCAAGAGGCTCAAGGCAACA 4058
 QY 2210 nLysThrSerValLeuGlyGlyGlyLysArgGlyIleGluProValSerProProGluGly 2230
 Db 4059 CAAGAGCTCGGCTTGGTGTGTGTGAAGACGTAATTGAACCTGTGCTCCCAACGAGAGG 4118
 QY 2230 yMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGly 2250
 Db 4119 CATGACGAGAGCCAGGCACTCCGAGAGTCTGTGTAACCCGCTGCTGACCGGATGGGGA 4178
 QY 2250 uGlnThrGluProSerArgMetGlySerLysSerProGlyLysThrSerGlnProProAl 2270
 Db 4179 ACAGACGAGGCCAGAGAGATGGGCTCCAAAGTCTCAAGGCAACACGAGCCGCGCAGC 4238

QY 2270 aPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysGlyGluIle 2290
 Db 4239 CTTCTTACAGCAAGCTGACCGAGAGCACTCCGCTATGTCAAGTCAAGAGAACAGAT 4298
 QY 2290 sAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGly 2310
 Db 4299 CAACAAAGAGTGAACACCCACCAACCGGATAGCTGTAATTCATATACGCAAGCTGG 4358
 QY 2310 yThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGly 2330
 Db 4359 GACGAGATCTTCAATATAGCCCGCATACCGGAACAGGCTTATAGACTTATAGAACCA 4418
 QY 2330 nAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGlyAlaIleIleArgValAlaLe 2350
 Db 4419 GCGGCTGACAGAACTGCGCAGACCAACATGGGGCTGGAGGCCAATTAAGAACGACT 4478
 QY 2350 uMetGlyLysTyrAspGlnThrGluGluSerProProLeuSerAlaAsnAlaPheAsnPr 2370
 Db 4479 CATGGGTAAATATGACCGAGGGAAGATCCCGCGCTCAAGCCCATGCTTTTAACCC 4538
 QY 2370 oLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSe 2390
 Db 4539 TCTGAATGCGCACTGCGCAGCTGCGCGCTATGCTATGCTGCTGACGAGCGAG 4598
 QY 2390 rAspHisThrLeuThrSerProGlyGlyGlyGlyValLysValSerGlyArgProSe 2410
 Db 4599 TGACCACACACTCACCTCGCGAGTGGCGCGGAGGCAAGGCTCTGCGAGACCCAG 4658
 QY 2410 rSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerVa 2430
 Db 4659 CAGCGAAGAAAGCAAGTCCCGGCGCCCGGCTTGGAATGGGAGACCGGCACCTCTGT 4718
 QY 2430 lSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTr 2450
 Db 4719 CTCTCAATGCACTGAGAGGAGACAGTGAACCGCGGAGCGGCTACCAACCGCGTGTG 4778
 QY 2450 pGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetLr 2470
 Db 4779 GGAGAGCAGGCCCTGCTCGCAGTTCACAGCCATTCCTTCAACCCCTTATCATGTGG 4838
 QY 2470 gLeuGlnAlaGlyValMetAlaSerProProProProGlyLysProAlaGlySerGlyPr 2490
 Db 4839 GCTGCAAGCGGGGTGATATGCTTCCACCCCAACCGGAGCTCCCGCGGAGCGGGCC 4898
 QY 2490 oLeuAlaGlyProHisHisAlaIleArgAspGluGluProLysProLeuLeuCysSerGlnTy 2510
 Db 4899 CTTGCTGCGCCCGCCACCAAGCTGGGAGCAGAGAGCCCAAGCCACTGCTGCTCGAGTA 4958
 QY 2510 rGluThrLeuSerAspSerGlu 2517
 Db 4959 CGAGACACTCTCCGACAGCGAG 4980

RESULT 9
 S83390 2930 bp mRNA linear PRI 28-MAR-1997
 LOCUS T3 receptor-associated cofactor-1 [human, fetal liver, mRNA, 2930
 DEFINITION nt].
 ACCESSION S83390
 VERSION S83390.1 GI:1911769
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 2930)
 Sander, S. and Privatky, M.L.
 Identification of TRACs (T3 receptor-associated cofactors), a
 family of cofactors that associate with, and modulate the activity
 of, nuclear hormone receptors
 Mol. Endocrinol. 10 (7), 813-825 (1996)
 JOURNAL MEDLINE
 PUBMED 86408715
 8813722
 Genbank staff at the National Library of Medicine created this
 REMARK

QY 2112 SerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValIysGlyHisGlnArg 2131
 DB 1683 ACCCAAGCCTCGTCCAGAGCCGCGTCTCCAGACCGGCCAGGGGTCAAGGTACACCGCGG 1742
 QY 2132 ValValThrLeuAlaGlnHisIleSerGlyValIleThrGlnAspTyrThrArgHis 2151
 DB 1743 GTGGTACCTCGGCCAGACATCACTAGTCACTACACAGACTACACCCGGCACAC 1802
 QY 2152 ProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyValaSerCys 2171
 DB 1803 CCACAGCAGCTCAGCGCACCTCTCGCCGCCCTCTTACTCTCTCTGGGGCCAGCTGC 1862
 QY 2172 ProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProAspHisGly 2191
 DB 1863 CCGGTCTGACCTCGCGCCGCCACCACTGACCTTACTCTCGCCGCCAGCATGTGT 1922
 QY 2192 AlaProAlaArgGlySerProHisSerGlyValGlyValArgSerProGluProAsnLys 2211
 DB 1923 GCGCCGGCCCGTGGTCCCCCAGACGAGGGGGCMAAGGTCTCCAGAGCCAAACAG 1982
 QY 2212 ThrSerValLeuGlyGlyGlyValAspGlyIleGluProValSerProGluGlyMet 2231
 DB 1983 ACCTCGCTTGGGTGGTGTGTGAGACGGTATTGAACCTGTGTCCACCGAGGGCATG 2042
 QY 2232 ThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGln 2251
 DB 2043 ACGAGACCGAGGCACTCCCGAGTGTGTGTACCTCTGTAACCGGAGTGGGGAACAG 2102
 QY 2252 ThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProAlaPhe 2271
 DB 2103 ACGAGACCGAGAGTGGGTCTCCAACTCCAGGACACACGACGCGCCAGCTTTC 2162
 QY 2272 PheSerLysLeuThrGlnSerAsnSerAlaMetValLysSerLysGlnGluIleAsn 2291
 DB 2163 TTCACGACACTGACCGAGACCACTCCGCTCATGTCAAGTCCAGAGACCAAGATCAAC 2222
 QY 2292 LysLysLeuAsnThrHisAsnArgAsnGluProGlyTyrAsnLysSerGlnProGlyThr 2311
 DB 2223 AAGAAGCTCAACACCCACAACCGAATGAGCTGATCAATCAATACGACCTGGGAGC 2282
 QY 2312 GluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAla 2331
 DB 2283 GAGATCTTCAATAATGCCGCCATCACCGGAAACGGGCTTATGACTATGAAGCCAGGGG 2342
 QY 2332 ValGlnGluHisIleAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMet 2351
 DB 2343 GTGCAGAAACATGCCAGACCAACATGGGGCTGAGAGCCATTAATTGAAAGGCACTCATG 2402
 QY 2352 GlyLysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeu 2371
 DB 2402 ----- 2402
 QY 2372 AsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAsp 2391
 DB 2402 ----- 2402
 QY 2392 HisThrLeuThrSerProGlyGlyGlyLysAlaLysValSerGlyArgProSerSer 2411
 DB 2403 -----GTTGGCGGCGGAGAGCCAAAGTCTCTGGCAGACCCAGCAAC 2444
 QY 2412 ArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSer 2431
 DB 2445 CGAAAGCCCAAGTCCCGGCGCCGCGCTGTGGTGTGGGAGCGGCACTCTGTCTCC 2504
 QY 2432 SerValHisSerGlyGlyAspCysAsnAlaGlyArgThrProLeuThrAsnArgValTyrGlu 2451
 DB 2505 TCAGTCACTCGAGGAGGAGCTGCAACCGCGGACCGCTCTCAACACCGCGTGGAG 2564
 QY 2452 AspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLysLeu 2471
 DB 2565 GACAGGCCCTCGTCCGACAGTTCCAGCCATTCCTCTTCAACCCCTGATCATGCGGCTG 2624
 QY 2472 GlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeu 2491

DB 2625 CAGGCGGGTGTCAATGCTTCCCAACCCCGGCGCTCCCGCGGAGCGGCCCTTC 2684
 QY 2492 AlaGlyProHisIleAlaTyrAspGluGluProLysProLeuLeuCysSerGlnTyrGlu 2511
 DB 2685 GCTGCGGCCACACAGCTCTGGGAGCAGAGGCCCAAGCCACTGTCTGTCTCGAGTACAG 2744
 QY 2512 ThrLeuSerAspSerGlu 2517
 DB 2745 ACACCTCTCCAGACGCGAG 2762
 RESULT 10
 AB028970 7949 bp mRNA linear PRI 02-APR-2003
 LOCUS Homo sapiens mRNA for KIAA1047 protein, partial cds.
 AB028970
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS
 TITLE
 Kikuno, R., Nagase, T., Ishikawa, K., Hirose, M., Miyajima, N.,
 Tanaka, A., Kotani, H., Nomura, N., and Ohara, O.,
 Prediction of the coding sequences of unidentified human genes.
 XIV. The complete sequences of 100 new cDNA clones from brain which
 code for large proteins in vitro
 DNA Res. 6 (3), 197-205 (1999)
 MEDLINE
 PUBMED
 99397452
 10470851
 2 (bases 1 to 7949)
 Ohara, O., Nagase, T., and Kikuno, R.
 Direct Submission
 JOURNAL
 COMMENT
 On Mar 31, 2003 this sequence version replaced gi:5689430.
 FEATURES
 source
 1..7949
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="17"
 /clone="h01221s1"
 /tissue_type="Brain"
 /note="vector: pBC SK plus"
 1..7949
 /gene="KIAA1047"
 <12..7058
 /note="start codon is not identified.
 This sequence was obtained by subcloning of the DNA
 fragments derived from two cDNA clones (1 - 1691 was
 derived from h01221 and 1692 - 7949 was derived from
 h02672)."
 h02672)."
 /codon_start=1
 /product="KIAA1047 protein"
 /protein_id="BA82999.2"
 /db_xref="GI:29421188"
 /translation="SHSLVKTSLIMSSGYPVNGAFSTEQSRYPHISVOYTPNTR
 HQQERAVPDYSSHLEVOASQLQQQQQQQLRRPSSLSEPHPSDPRPQERTSYEP
 FHPGSPVHDLSLEKRPLEQSDSHQVSAVLPVHPLEGLASADAKQDAF
 GSKHAPSPTSGQPCGNDONASPSKSKSELISGMDVDRILAVEQDILKKKKQ
 QLEEEAARPPPEKPVSPPYEQKRSIVQIITDNRKKAREAHKIFSGIGPKVPLP
 INQPSDVTYHNIKTNQVNRKGLIFTKRNRHAKQREOKTCQYTDQLEMEWEKVD
 RIENNRRAKESKTRERKQPEPIRQRQEQERFQVGGAGLSATLASSEHS
 EIIIDLSQENNEKQROLSVIPMMFPAQGRVYFIMNGIMEDPMKRYKORPMNV
 WTDHEKEIFKQFIQHPKNQGLIASYLRKSYDCLVLYYLTKENYKALVRRYK
 RGRNOQIARSPQEKVEKEDEKAKETKEKEEKKDEEDKESDKENTKDKID
 GTAETPEREQATPRGRTKANSQGRKRKIRTSRWNEAASAAATTEEPPLPP
 PPEPISTPERVETSRWTEEMEVAKKGLVHGKRWMAIAIKMTGTSKSEACKNFYNYKR


```

Db      1626 AAAAGAGAAAGAAAGAAATGAGAGAAAAAGAAAGAAAGAAATCCAAAGAAAT 1685
Qy      541 AApLySGlAAspLeuLeuLySGlulYsThrAAspAspThrSerGlyLuAspAAspGlu 560
Db      1686 ACCAAGAA-----AAGACAAAGATAGATGATACAGA---GAAGAACTGAGAA 1733
Qy      561 LySGlualAvalAserLySGlYarGlySthralAAsnSerGlnGlyarGlySGly 580
Db      1734 AAGAGACAAAGCAACACCCCGGGGCGAAAGACTGCACACAGTCAGGGCGCGCTAAGGCG 1793
Qy      581 ArgLleThrArgSerMetAlaAsnGluAlaAsnSerGluLualAlleThrProGlnGln 600
Db      1794 CGATTCACCAAGCTCATGACAAACGAAGCTGACAGCTGCTCAGCCGACGGCGCT 1853
Qy      601 Ser-----AlaGluLeuAlaSerMetGluLeuAn 610
Db      1854 ACTGAAGAGCCCCCACCACCTCTGCCACCGCCACCGAAACCCATTCTACAGAGCCTGTG 1913
Qy      611 GluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLySGlyLeuLeuGluHs 630
Db      1914 GAGACCTCTCGATGACAGAAAGAAATGGAAGTTGCTAAAGAGTCTAAGAGAATCAT 1973
Qy      631 GlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLys 650
Db      1974 GGTCGTACTGGGCAACAAATGCTAAATGTGTGGAACGAAGAACTCAATGTATAA 2033
Qy      651 AAspPheTyPheAsnTyrlYsLysArgGluAsnLeuAspGluIleLeuGlnGlnHisLys 670
Db      2034 AACTTATATTATTAACCTRTAAAGCGCACACATCTTGACCACTCTTACAGCGCATAA 2093
Qy      671 LeuLysMetGluLysGluArgAenAlaArgArgLysLysLysLysAlaProAlaAlaAla 690
Db      2094 CAGAAAACTTCAGAAACCTCGTGAAGAGAGATGTGTCTCAATGTGAAGTGTGCT 2153
Qy      691 SerGluGluAlaAlaAspProProValValGluAspGluGluMetGluLysSerGlyVal 710
Db      2154 TCACACTGTTCTGCT-----CAAGAGATGAAGATATTGAAGCTTC----- 2195
Qy      711 SerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGlu 730
Db      2196 -----AATGAAAGAAAGAAATCCAGAAAGACGAAAGT----- 2228
Qy      731 ValProArgGlyLysCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSer 750
Db      2229 -----GCAGAAATATAGTTCTGTATACAGAAAGT 2255
Qy      751 IleProSerProHis---ThrGluAlaAlaLysAspThrGlyLysAsnGlyProLysPro 769
Db      2256 GCTCTCTTCTCTTCCACCACTTGAAGCTGTCAAG----- 2288
Qy      770 ProAlaThrLeuGlyAlaAspGlyProProProGlyProProThrProProAlaArgThr 789
Db      2289 -----CCACGAGAGACAGCTCTGAAAAATGCTACT 2318
Qy      790 SerArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProPro 809
Db      2319 TCTCGAGAAACACAGAACCTCGCGTTGAGCTTGAG-----CCACACGACGAA 2366
Qy      810 ProAlaPro---ProSerProSer---AlaProProProValValProLysGluLys 827
Db      2367 ACTGACCCAGTACATCTCTCTTACAGATTCCAAAGTCAAAACCACTGAAGATGA 2426
Qy      828 GluGluGluThrAlaAlaAlaProProVal----- 837
Db      2427 AGCTGTGAAGACCAAGGTGAATGACAGTACGTCTGAGACACAGACAGCATGATGA 2486
Qy      838 -----GluGluGlyGluGluGluLysProProAlaAlaGluGlu 850
Db      2487 GATCAGCAGGAGACAGTGTCTGAAGAGGTCTGTGTGTATCCCCACCGCTACCAAA 2546
Qy      851 LeuAlaValAspThrGlyLysAlaGluGluProValLysSerGlySthrGluLualA 870

```

```

Db      2547 -----GCTGACTCTGTGACCTTGATGAGCGTCCAGAAACCATGCTATAAGT 2600
Qy      871 GluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAla 890
Db      2601 GAAAGGTAAATATACAAAGAAAGAGACTTGTGATGAGCC----- 2639
Qy      891 LeuLysAlaGluLysGluGluGlyLysGlyAlaGlyAlaThrThrAlaLysSerSerGly 910
Db      2640 -----AGTGAAGAGGTGAACCTAGAGATGAAGATTTGGTGTAGCTCAGCAATTAAT 2693
Qy      911 Ala-----ProGln---AspSerAspSerSerAlaThrCysSerAlaAspGlu 925
Db      2694 GCCCAAGGCCCGGAGCCCGCATGACACATATGATTCAGTCCGACGCGCATGTATGAG 2753
Qy      926 ValAspGluAlaGluGlyLysLysAspLysAsnArgLeuLeuSerPro-----ArgPro 942
Db      2754 -----GATGTGATGTGACAGCCAGAGAGGACAGAGATGTTCTGTATGACATCAAGCCT 2807
Qy      943 SerLeuLeuThrProThrGlyAspProAlaGluAlaAsnAlaSerProGlnLys-----Pro 960
Db      2808 TCACGTGTTAAACCCCACTGATCTTATCTGTC---TCATCTCCCTTAAACCAATCCA 2864
Qy      961 LeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIle----- 976
Db      2865 CTGATCTGCCACAGCTTACAGATGAGCTGCTGTATCCACCAATGATTCCTGCACC 2924
Qy      977 -----GlnVal 978
Db      2925 CCATGTAACATACCAATTTGAACCCCGATGAGCGGCTATGCTCTTACAGCAGCACATTT 2984
Qy      979 ThrLysValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAla 998
Db      2985 AAAGCAATGTCATGAGTCAGCACTCTGAG----- 3014
Qy      999 ProProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSer 1018
Db      3015 -----GAGCAGCGGACAGACAGAACAGATAGATTTGGATGTGAAGTTCTACAAGT 3068
Qy      1019 ProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaAlaAlaGlu 1038
Db      3069 CCATGTGGACATCCCAAGAGTCCA-----AACAGAGAG----- 3101
Qy      1039 AlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValPro 1058
Db      3102 -----TGG----- 3104
Qy      1059 ProArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyAla 1078
Db      3105 -----GAGTCTCTTACGCTGTCCACAT-----CAAGTGAATTAATCTC 3146
Qy      1079 ProProGlyHisAspLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArg 1098
Db      3147 CCGAAGAGGCGTTCCGCTTCG-----ACAATCGAACCAACAGGCCACCG 3191
Qy      1099 ProProThrIleSerAsnProProProLeuIleSerSerAlaLysHisAspSerValLeu 1118
Db      3192 CCCCTCTCATCCCGTCATCCAAACACACAGCGCTTGCAAAA---CCATCTTTTATA 3248
Qy      1119 GluArgGlnIleGlyAlaIleSerGlnGly-----MetSerValGlnLeu 1133
Db      3249 -----ATGGAGAGCTCATCTCAGAGGAACACAGGACATTAATTGACTTCTATAT 3302
Qy      1134 HisValProTySerGlnHisAlaLysAlaPro---ValGlyProValAlaThrMetGlyLeu 1152
Db      3303 CAGGCTTCTTACATCAAGAAACACCAAGCGTCAAGTGGATCTATCTCTGTGACTG 3362
Qy      1153 ProLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLeu 1172
Db      3363 CCACGGCAACAGAAATCTCCAAATCACTACTTGGCCCTTACATCAAGCAGGAAGATTT 3422
Qy      1173 SerProArgGlyGlnAlaGlyProProGluSerLeuGlyValAlaProThrAlaGlnGluAla 1192
Db      3423 TCTCCCGAAGCCAAACTCAACACTGAGGCTCTGTGTGTC---AGGCGCCAACTGAA 3479

```

[illegible][illegible]

```

Db      5400 CAGAGCCCAAGTGTTCCTCAAGAACCAATGGAACCAAGTATATCAACCTTTGGATGCCA 5459
Qy      1893 SerLysProThrValLeuArgSerThrSerThrSerProValArgProAlaIaIaThr 1912
Db      5460 ACT-----GCTCAGCTTACGA-----ATCATGCCA----- 5483
Qy      1913 PheProProAlaThrHisCysPProLeuGlyGlyThrLeuAspGlyValTyrProThrLeu 1932
Db      5484 -----CTGCCTGCTGGGGGC----- 5498
Qy      1933 MetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProArg 1952
Db      5498 ----- 5498
Qy      1953 AlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAla 1972
Db      5499 -----CTTCAATAAGCCAGAGCCTG----- 5519
Qy      1973 SerSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisVala 1992
Db      5520 -----CCAGCCTCCCGTTATCAAC 5537
Qy      1993 ThrIleAlaArgThrProAlaLysAsnLeuAlaProHisIleAlaSerProAspProPro 2012
Db      5538 ACTGCTGCGAGATGCCCTGCTCTGTG----- 5567
Qy      2013 AlaProProAlaSerAlaSerAspProHisArgGlyLysThrGlnSerLysProPheSer 2032
Db      5568 -----GATGCTGCAGCTTCTGCACCCCAAGATGATGTGTCCAAACAAAGAGATATCT 5621
Qy      2033 IleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerThrSerProGluGly 2052
Db      5622 TCTCAGAGGTATGAAACACCTAGC-----GATGCT 5651
Qy      2053 ValGluProValSerProValSerProValSerProLeuThrHisAspLysGlyLeuProLys 2072
Db      5652 ATTGAGGTATAGTCTCTGCACAGTCACTGCAGCCCAACCCAGAGAACTGCAGACCTAT 5711
Qy      2073 HisLeuGluGluLeuAspLysSerHisLeuGluGlyGlyLeuArgProLysGlnProGly 2092
Db      5712 CAGCCAGAGGTTTAAAGCAAAATCAAGCGGAAAGATCTTCACAGACAAATGAGGA 5771
Qy      2093 ProValLysLeuGlyGlyLysAlaAlaHisLeuProHisLeuArgProLeuProGluSer 2112
Db      5772 CCA-----TTACATCATATGACCA-----CAG 5795
Qy      2113 GlnProSerSerSerProLeuLeuGlnThrAlaProGly----- 2125
Db      5796 CAGGAATCAACATCTCCCAACAAACAGCTGCCCTTTCTTCACAGGCAAGAGGAATGGG 5855
Qy      2126 ---ValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThr 2144
Db      5856 CAAAGTCCCAAGAACCATGGCTGATCAACATCTGCTGATCAACATCTGTCAAAATATACAA 5915
Qy      2145 GlnAspTyrThrArgHisHis-----ProGlnGlnLeuSerAlaProLeu 2159
Db      5916 CAAAGATTTCATGATGAATCAAGTTTCTCGCAGACCTCCAGAG-----CCTCCT 5966
Qy      2160 ProAlaProLeuTyrSerPheProGlyAla-----SerCysProValLeuAspLeuArg 2177
Db      5967 ACTTTCATCATTCAGAACTCACTTCTGCTTGGTATCAACACTGTG-----AGG 6017
Qy      2178 ArgProProSerAspLeuTyrLeuProProProAsp-----HisGlyAla 2192
Db      6018 ACTAAACATCAAAACCGTTACAGCCCAAGATCCAGAGCTCAGCTGTCCATCATCAAAAG 6077
Qy      2193 Pro--AlaArgGlySerProHis-----SerGluGlyLysArgSer 2206
Db      6078 CCAGGTTCAAGGGGTCTCTCAGAAATCTTGTGCAAAATCCAGGGGAAGTAGGCTTGA 6137
Qy      2207 ProGluProAsnLysThrSerValLeuGlyGlyGlyLysAspGlyIleGluProValSer 2226

```

```

Db      6138 AAATCCCAAGAGAGATCACGTC-----TCTTGGAGACCTTACAGGCCATCTCC 6188
Qy      2227 ProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyr 2246
Db      6189 CCACCCAG-----GTTCCGGTTGGATGAGAAACAGACAGCTTGCTGCTGTCT 6242
Qy      2247 ArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSer 2266
Db      6243 CAGAGGGGGCGAGAGCCTGCAGAGCAAGAAATGATGCCGCTCACAGGGAGTATTAAGC 6302
Qy      2267 GlnProProAlaPhePheSerLysLeuThrGlySerAsnSerAlaMetValLysSerLys 2286
Db      6303 TACTTGCCTTCATTTCTTCCACCAAGCTT---GAAATATCATCAACCATGGTTAAATCAAG 6359
Qy      2287 LysGlnGluIleAsnLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIle 2306
Db      6360 AAGCAGAGATTTTTCGTAAAGTTGAATCTCTGTGGAGGTGACCTGATATGGCAGCT 6419
Qy      2307 SerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThr 2326
Db      6420 GCTCAGCCAGAACTGAGATCTTTAATCTGCCAGCAATTAACGTACAGGCTCAGTTAGC 6479
Qy      2327 TyrArgSerGlnAlaValGlnGluHisIleSerThrAsnMetGlyLeuGluAlaIleIle 2346
Db      6480 TCTAAGGCCATTTCTTTGCTGATCTCTGCACT---AATCTTGGCTGGAAGACATTAAC 6536
Qy      2347 ArgLysAlaLeuMetGlyLysTyrAspGlnTyrGluGlu-----Ser 2360
Db      6537 ACGAAGGCTCTCATGGAAAGCTTGATGACAAAGTTGAGATCATGAGAGTTGATGATGCC 6596
Qy      2361 ProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaIle 2380
Db      6597 CAGCCTATGGAGTA-----GTGCTGTACTGTCACAACTCAGTT----- 6638
Qy      2381 MetProIleThrAlaAlaAspGlyLysSerAspHisIleThrSerProGlyGlyGly 2400
Db      6639 -----GTACCAAGTGTGACACAGAAAGAGAAAGGAGCACTACCTATTCAGGA 6692
Qy      2401 Gly-----LysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAla 2418
Db      6693 GGAGTTTGCAACCAAAAGCTGATCAGCAAGTCAACAGCAGGAAATCTAAAGTCTCTATA 6752
Qy      2419 Pro-----GlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGlu 2436
Db      6753 CTTGGGCAAGGCTTACTTAGAACGGAACGCCCTCTTCACTCTCTGTATACATTCAGAA 6812
Qy      2437 GlyAspCysAsnArgArgThrProLeuThrAsnArgValTyrGluAspArgProSerSer 2456
Db      6813 GGGATTTACATAGGACAGCCCA-----GGGTGGCCTGGGAAACAGGCCCTTTCA 6866
Qy      2457 AlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMet 2476
Db      6867 AAGAGCTCAACTCAGTTTCTTATTAACCTCTGACATATCGGATG-----CTCAGC 6917
Qy      2477 AlaSerProProProProGlyLeuProAlaGlySerGly---ProLeuAlaGlyProHis 2495
Db      6918 ACTACTCCCAACAACCGATTGTCATGTGCTCCCTGCGGTGAACCAACACACTCTCTAC 6977
Qy      2496 -----HisAlaTyrAspGluGluProLysProLeuLeuCysSerGlnTyrGluThr 2512
Db      6978 CAACGAAACAGATCTGGAGCGAAGACCTGCCCACTGCTCTCAGACAGTACAGAGCC 7037
Qy      2513 LeuSerAspSerGlu 2517
Db      7038 CTGTGGATATGAT 7052

```

RESULT 11
 MMU35312 780 bp mRNA linear ROD 19-OCT-1996
 LOCUS
 DEFINITION Mus musculus nuclear receptor co-repressor mRNA, complete cds.
 U35312
 ACCESSION
 U35312.1 GI:1022717
 KEYWORDS

SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 7780)
AUTHORS Horlein,A.J., Naar,A.M., Heinzel,T., Torchia,J., Gloss,B., Kurokawa,R., Ryan,A., Kamei,Y., Soderstrom,M., Glass,C.K. and Rosenfeld,M.G.
TITLE Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor
JOURNAL Nature 377 (6548), 397-404 (1995)
MEDLINE 96008639
PUBMED 7566114
REFERENCES 2 (bases 1 to 7780)
AUTHORS Horlein,A.J., Naar,A.M., Heinzel,T., Torchia,J., Gloss,B., Kurokawa,R., Kamei,Y., Ryan,A., Glass,C.K. and Rosenfeld,M.G.
TITLE Direct Submission
SUBMITTED (02-SEP-1995) Andreas J. Horlein, School of Medicine, UCSD, 9500 Gilman Drive, La Jolla, CA 92093-0648, USA
LOCATION/Qualifiers
1. 7780
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/issue_type="pituitary"
117. 7478
/note="N-COR"
/codon_start=1
/product="nuclear receptor co-repressor"
/protein_id="AA017125.1"
/db_xref="GI:1022718"
/translation="MSSSGYPNQAGAFSTEQSRSHVQYTPSPAAHQEFAVPDRA
SHLEYSQASQLQQQQQQLRRRPSLSPHSGSDRDRRSGAGVDFHFGSPVND
SLESKRPLEQVDSHQRQISAVLPVHLPGLSGRSNAKKDAPAGVHGAHPSPL
SGPCPDQNAKSKSKEELISMDRYDEIAKVEQILKAKKQQLKEEAKKPE
PERKVPPEVQKHRSIVQIYDENRKAKEAHKIFEGLPKVELPLVNPSTKYX
ENIKTVQWNRKLIFFKRNHAKKOEQICQYDOLAMWKKVDRIENRRKX
ESTKREYKQPEIRKQEQERFORCGAGLSATILASHEI SEI IDGSEON
NEKOROLSVIPMMFPAEORVAFIMNGMEDPMKVFQROPMVMTDHEKIFPD
KFIQHPKNGELIASYLERKSPVPCVLYVLTKNENKALVRNRYGKRRNOIAR
SOEKVEKEDAKTEKKEBEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDE
VTRGRKTASQGRGKRVTRSTSEAAANAANAATEEPPLPEPPELSTIEPVT
SRWTEEMEVAKGLVEHGNMAAIAWVQTKESBAQCKNYFYKRNHNDNLQOHK
OKASRRPREEDVQCEVASVTAQDEDI EASNEEENDEBGEANSDDTSASBP
SPYBAKSSSEDSSENAASRGTEPVAELAKTDDPAPCASPVPTTKPERSVRA
VDSASAEETAPNDVHEGCGSSVLDPAPTKADSDVPMQVVENTASKEGAK
ERLESTSEKTEKDEDDVVAQIEREPEDSDSDSATGCGADGVGEPEKRVFPMD
AKSLILTPGSLILISSPIKPNLIDLQLOHRAAVIPWVSCTCNPIPIGPVSGALY
ORHILHLEBALBEQROROBOVDLECRSSTSPSTSKSPREMEVLOPAHOVITNL
PEGVRLPTTPRPPLPSSKTVAEKSFPISGSGISQTPGYLSHNOVAQOE
APKVSQISLGLPROESTKAPLTYIKQEEFSPGSONSGPGLVLRAGHEVAGT
AGAVQGSITRGTPAKTISVETISLARGSTTQCTPALPQAGITTEALVKGVPVMPLE
ESSEKVEBEBAKGNHIVKSGHILSYDNINAKBGTSPRTAHMSLKRSYEAIVE
GSIKQGSMEKESPVSAVLEGLICRALPGRSPHDLKERTVLSIGOGTPRAVAEPE
DGAKPKQIKREBPPIRAPEGATITKGPYGIITTIKEMGSIHEIPRODLTQESKRT
PEVQGTROIKEGSIQGTPIKPDNNSGQALIKHVSGLTSGSKLPRGLMEIVPNTI
KVERGKVEDVKADEPRARATSVSSGSEVLSTLHEAKALSGLVNGLVRSARBP
VYQNTISGSPMMNRTSDVSSSKASHKSTLTPQKRSITAKSPREVDVDTVS
PFPPIHRSAGAEVTRSHLTPHDPAPFRALDPAAVLYLOQLSTPEVPSQYQY
AMENTQOTLINDYITSQOMQVNLPRVTRLSPREQLPAGITLIDLTQVPLTI
LVHAGGTSTPMDRIITYIPGTQVTPPPRYNAASLSPGPTLAAASAEERERER
EKREKEREKEREKEREKRIIAADLYLRGSGRPGRSGGVNRPSPVAFQETL
QOQSVFOGNGTSVITPLDPTAOLIMPLPSGGPSISQGLPSRNVNTADALVLD
AAASAPOMVSKTESKREKHALENLRSASAVSBOOLEKQLEKLEVERSGVCS
SALPSKAOQPHASVUYSEAKDKGPPPKSYEEELRTGTITTAANFDIVITTRQIA
SDDAKERSQSDSSSSUSHRVETASDIIVISPSAPPOKQPOAOPVIVAN
QANESITROVEGLAHYRQGESPSQOQSPQSPQSEMGQVPTHTLITLADIC
QIITDPANOVPSQASTFTQSPALSTPVRTKTSRSGVSESOQTYLHRRPRA
VBERNLVDSRGRSGKSPERSHISEPYPISSPOOAVHEKODMLLSQGVNPA
BOESDERSGSIISYLPSPFTKLESTSPMKYSKQEI PKLNSGGSDMAAAOQPT
IRPLPAVTTSAGVSSSHSPADPASNUGLEDTIRKALMSFPDKVDEHCVMSHPGI
MPSASTSVYTHSEARDBGSPPHAGVCPKLINKNSKSPSPGQSYLDTETSPS
SVSSVSEGDYHRTQPGMAWEDRSPSSGTQFPNPLITIRMLSSSTPTQIACAPSAIT

ORIGIN QAAPOONRIWEREPAPLISAOYETLSDSD"
Alignment Scores: 1.68e-76 Length: 7780
Pred. No.: 4187.00 Matches: 1104
Score: 53.94% Conservative: 347
Percent Similarity: 41.04% Mismatches: 798
Best Local Similarity: 31.68% Indels: 442
Query Match: 10 Gaps: 104
DB: 10
US-09-522-753-5 (1-2517) x MMJ35312 (1-7780)
Qy 16 GluProArgTyrProProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThr 35
Db 165 CAGAGTCGCTACCTTGGCACTGGTTCAGTACACCTTCCAGCGCCCTGCACACAG 224
Qy 36 AspvAlcIleuLeuGlnTyrGln-----HisHisSerArgTyrAlaSerHisIleu 53
Db 225 GAATTTCAGTCCCTGCTACCTGCTTTCATCTTGAAGTCAGTCAAGGCTGCAGCTC 284
Qy 54 SerProGlySerIleIleGlnProGlnArgArgProSerIleuLeuSerGluPheGln 73
Db 285 TTCACAGACGACGACGACGACGCTTGCAGACGCGCTTCCGATTCCAGATTTCAC 344
Qy 74 ProGlyAsnGluArgSerGlnGluLeuHisIleuArgProGluSerHisSerTyrLeuPro 93
Db 345 CCGGCTTCCGACACGCGCCCGAGAA-----AGAGAGAGTGAATACAGACACTTCACCG 398
Qy 94 GluLeuGlyIysSerGluMetGluPheIleGluSerIysArgProArgLeuGluLeu 113
Db 399 GGCCTTCCCGGCTGACCATGCTCGTGAATCCAGAGCGCTGCTGAGCAGCTT 458
Qy 114 -ProArgProLeu-----LeuArgProSerProLeuLeuAlaThrGlyGlnPr 129
Db 459 TCCGACTCCCATCTCAGCGCATCAGTCTGCTGCTCTCTTGTGTGCACACG-CTGCC 517
Qy 129 oAlaGly-----SerGluAspLeuThrIysAspArgSerLeuThrGlyIysLeuG 146
Db 518 AGAAGACGACGAGTCTTCTCCCAATGTAAGAAGATCCGCACTTGGATCAAACATGA 577
Qy 146 u--ProValSerProProSerProProHisThrAspProGluLeuGlnIleuValProPr 165
Db 578 AGCTCTTCTCTCTCTCTCTCTCTGAGCAGCATCGAGATGATCAGAACTCTCACCTTC 637
Qy 165 oAlaGluSerIysGluGlnIleuIleGlnAsnMetAspArgValAspArgGluIleThrMe 185
Db 638 AAAACGTCAAGAGAAAGAGCTGATACAGACATGATCGTGTGATCGAGAAATTGCGAA 697
Qy 185 tValGluGlnGlnIleSerIysLeuIysIysIysGlnGlnGlnIleuGlnIleuAlaI 205
Db 698 AGTGAACACGACATCTTAACTGMAAAGAAAGACGACACCTGMAAGAAAGCTGC 757
Qy 205 alyProProGluProGluIlyeProValSerProProtoiIeGluSerHisIArgSe 225
Db 758 TAAACCCCGACGACCTGAGAAAGCTGTGCTCTCTCCGCTGAGAGAACCGAAG 817
Qy 225 tLeuValGlnIleIleTyrAspGluAsnArgIlyIysValIleGlnAlaIHisIArgIle 245
Db 818 TATTTGCCAAATCATTAATTAAGACAGAAATGGAAGAAAGAGAAAGCTCATTAATATT 877
Qy 245 uGluGlyLeuGlyProGlnValGlnIleuLeuProLeuTyrAsnGlnProSerAspThrArg 265
Db 878 TGAAGCTCTTGCCCAAAAGTTGAATCGCGCTCTACACACGCGCTCAGATACCAAGT 937
Qy 265 nTyrHisGluAsnIleIysIleAsnGlnAlaMetArgIlyIysLeuIleuTyrPheIy 285
Db 938 GTACCAACGAGACATCAAGACAAACAGAGTGAAGAAAGAACTCATTTATTTTAA 997
Qy 285 sAlaGluAsnHisIleAlaArgIlyGlnIleTyrIysGlnIlySerGlnArgTyrAspGln 305
Db 998 AAGAAGAAATCATGACAGAAACAAAGAGAAACAAAGAAATCTGCAACGTTATGATCAGCT 1057

QY 305 uMecGluAlaLeuGluLysLysValGluArgGluLeuAsnProArgArgArgAlaLys 325
DB 1058 CATGGAGCATGGAGAAAAAGTGCAGCAATAGAAAAATTAATCCCGAGGAAACCTAA 1117
QY 325 GGUUserLysValArgGluLysLysValArgGluLysGluLysGluLysGluLysGluLys 345
DB 1118 AAAAAAGCAAAACAGGAAATACATATGAAAAAGCATTCCAGAAATTCGAAACAAAGAA 1177
QY 345 uLeuGluGluArgMetGlnSerArgValGluArgGluLysGluLysSerMetSerAla 365
DB 1178 ACAGCAAGAAAGATTTCAG---CGAGTGGTGCAGAGGGAGCGTGGCTTTCAGCCACCTA 1234
QY 365 AlaArgSerGluLysGluValSerGluLysLeuArgGluLysGluLysGluLysGluLys 385
DB 1235 TCCTAGAGTGCAGTCAATGAGATTCGAAATTAATGATGCTTTCGAAACAGAGAAATA 1294
QY 385 uGluLysGluMetArgGluLysValAlaLeuProMetLeuLysArgAlaAspGlu 405
DB 1295 TGAGAAAGCAAAATGCGTACGCTTCTGTGATCCACTATGATGTTGATGCGAAACAAAG 1354
QY 405 nArgGluLysPheLeuAsnMetAsnGluLeuMetAlaAspProMetLysValLysLysAs 425
DB 1355 AAGGTCATATTCATATGATGATGCGTGCATGAGCATCCATGAGGTTTATAAAG 1414
QY 425 nArgGluValMetAsnMetTrpSerGluGluLysGluLysPheArgGluLysPhe 445
DB 1415 CAGACAGTTTATGATGTTGAGCTGACATGCAATGAAAAAGAGATCTTTAAGACAAAGTTTAT 1474
QY 445 rGlnHisProLysAsnPheGluLysLeuLysPheLeuGluArgLysThrValAlaGlu 465
DB 1475 CCGAGATCCAAAAACCTTGAGCTATGATCTCTATTTGAAAGAGAGAGGTTCTCTGA 1534
QY 465 uCyValLeuLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 485
DB 1535 TTGTGTTTATATTAATTAATTAACAAGAAAAATGGAATTAAGCCCTGCGTGAAG 1594
QY 485 gSerLys---ArgArgArgGluLysSerGluGluGluGluGluGluGluGluGluGlu 504
DB 1595 GAATTAATGAAAAACGACAGGCGAAGATCGAGATTCCTCCCTCAACAAAGAAAA 1654
QY 504 nGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 524
DB 1655 AGTAGAAGAAAG-----GAAAGAGATTAAGCAAAAAAAC 1690
QY 524 uLysGluLysGluAlaGluLys-----GluGluGluLysProGluValGluAsnAspLys 542
DB 1691 AGAGAAAAAGAAAGAAAGAAAGATGATGAAGAAAAAAGATAGAAAGAAACCTAA 1750
QY 542 GGUUAspLeuLeuLysGluLysThrAspPheTrpSerGluLysAsnAspGluLysGlu 562
DB 1751 AGAAACAAACCAAGAAAGACAGAAACGAAACGACAGCAAGAAACCTGAACAAAGAA 1810
QY 562 uAlaValAlaSerLysGluArgLysThrAlaAsnSerGluLysArgArgLysGluLys 582
DB 1811 GAGGCTACTCCAGAGGGGAGAAAGCTGCTAACAGCAAGGCGGAGGAGGGCGGT 1870
QY 582 eThrArgSerMetAlaAsnGlu-----AlaAsnSerGluGluAlaIleThrProGlu 599
DB 1871 CACCAAGTGCATGACAGTAGAGCTGACGCTCAATGCTGCGTGCAGGCCCTAGGA 1930
QY 599 nGlnSerAlaGluLeu-----AlaSerMetGluLeuAsnGluSer 613
DB 1931 GCCCCCGCACCCCTCCGACACACACAGCCCATTTCTACAGAACTGTTAGACTTC 1990
QY 613 rArgTrpThrGluGluGluMetGluThrAlaLysLysGluLysGluGluGluGluGluGlu 633
DB 1991 AAGATGACAGAAAGAAAGAAAGTGTCTAAAAAGCGCTGTGAACATGTGTGTA 2050
QY 633 nTrpSerAlaIleAlaArgMetValGluSerLysThrValSerGluCysLysAsnPheLys 653
DB 2051 CTGGGACGACATTGCTAAATGCTGGGAATAAAGTGAAGCCAGTGCAGAAAACTTCTA 2110
QY 653 rPheAsnTrpLysLysArgGluLysLeuAsnGluLysLysLysLysLysLysLysLys 673

DB 2111 TTTTAACCTATAAAGACGGCATTAATCTTGACAACTTTTGACGACAACTAAACAGAAAC 2170
QY 673 rGluLysGluArgAsnAlaArgArgLysLysLysLysLysLysLysLysLysLysLys 693
DB 2171 TTCACGAAACCCCGTAGAGGAAGATATCTCAGTGTGAAGTGTGCTTCCACTGT 2230
QY 693 uAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGluValSerGluLys 713
DB 2231 TTCCTCC-----CAAGAGATGCAATATTTGAAGCTCA-----AA 2266
QY 713 nGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGluAsnGluValProArg 733
DB 2267 TGAGAGGAAATCCAGAAAGATGTAAGCT----- 2297
QY 733 gGluGluCysSerGluProAlaThrValAsnAsnSerAspThrGluSerIleProSe 753
DB 2298 -----GCTGAAATATGTTCTGATACGAAAGTCTCTC 2332
QY 753 rProHis---ThrGluAlaAlaLys-----AspThrGluGluAsnGluProLysPhe 769
DB 2333 TCCTTCAACAGTTGAAGCTGCCAAGTCCAGTGAAGACAGAGTGAATAATGCT----- 2384
QY 769 oProAlaThrLeuGluAlaAspGluProProProGluProProThrProProArgTh 789
DB 2385 -----GC 2386
QY 789 rSerArgAlaProIleGluProThrProAlaSerGluAlaThrGluAlaProThrPro 809
DB 2387 TTCTGAGGAAACACCGAGCCGTGCTGAGCTGAGCCACACATGACCTTGCACCTG 2446
QY 809 oProAlaProProSerProSerAlaProProProValValProLysGluGluLysGlu 829
DB 2447 TGCATCTCCC-----TCTTCAGCAGTTCCAAACCAAAACAGCAAGAAAGGAAAGCT 2500
QY 829 uGlu-----ThrAlaAlaAlaProProValGluGluLysGlu 841
DB 2501 GAGAGCCAGGTGACCGACAGCGCCAGTGCAGAGCCGAGAGCCGATGAGCTAGACA 2560
QY 841 uGluGlu-----LysProProAlaAlaGluGluLysLys 852
DB 2561 TGAAGATGCGGTCGCGAGGCGAGTTCGTTCTGATCCACGAGCCCTACAA----- 2615
QY 852 aValAspThrGluLysAlaGluGluProValLysSerGluCysThrGluGluAlaGlu 872
DB 2616 -GCCGACTCCGTGACCCAGAAATGCAAGT---CCAGAAAAATACCTGCTTAAGAGTGA 2671
QY 872 uGluProAlaLysGluLysAspAlaGluAlaAlaGluAlaThrAlaGluLysLysLys 892
DB 2672 AGGAGATGCCAAAGAAAGACTTGAGAGCAC----- 2705
QY 892 aAlaGluLysLysGluGluLysGluLysArgAlaThrThrAlaLysSerSerGluValAsp 912
DB 2706 -AGTAGAAGACAGAGCGCTAAGATGAAGACGAGTGAAGTGAACAGATGAGAGGCC 2764
QY 912 oGlu-----AspSerAspSerSerAlaThrCysSerAlaAspGluValAspGlu 929
DB 2765 TGAGCACAGTACGACGAGCTCCAGTCCACTTGCACTGGGAGAGAGGTGTGATGG 2824
QY 929 aGluGluGluAspLysAsnArgLeu-----LeuSerProArgProSerLeuLeuThrPr 947
DB 2825 AGAG---CCAGAGAGGACAGAGAGTGTTCATGATGCAAAAGCTTCAITGTTAACTCC 2881
QY 947 oThrGluAspProArgAlaAsnAlaSerProGluLysPro-----LeuAspLeuLysGlu 965
DB 2882 TCCTGATCTATCTTAATC---TCAATCCCTTTAAACCAAACTATGATGCTGCACA 2938
QY 965 nLeuLysGluArgAlaAlaAlaIleProProIle----- 976
DB 2939 GCTTCAGCATGAGCTGTGTTATTCACCAATGTTCTTGACATCACTGATATATACC 2998
QY 977 -----GluValThrLysValHisGlu 983

Db 2999 AATTGGAAGCCCTTAAGTGGCTATGCTTTTACCAAGGACATTAAGGCATGCATGA 3058
 Qy 983 UPTGTAAGTGLuapAlaIaIaProthrlYsProAlaProProAlaProProProgl 1003
 Db 3059 GTCAAGCACTCTGTGAG-----GAGCAGCG 3082
 Qy 1003 naenleuGlProgluSerApAlaProglInglProglYsSerProArglYlYsSe 1023
 Db 3083 GCAGAGCAAGAAAGAGTGAAGCTTGAATGCAAGAGCTTACAAAGCCCATGACACACTTC 3142
 Qy 1023 rArGSeProAlaProProAlaAePlySGlualAphenAlaIaaglualaglInlylYleuPr 1043
 Db 3143 TAAGAGTCCA-----AACAGGAG----- 3161
 Qy 1043 oGlyArpProCysrPrThrsrGlyleuProPheProValProProArglYlVal11 1063
 Db 3162 -----TGG-----GAAATCT 3172
 Qy 1063 eLyAlaSerProhIaIaProArpProSerAlaPheSerTyAlaProProglYhIaPr 1083
 Db 3173 CCAAGCTGCTCCGAT-----CAAGTATTAACCTTCTGAAAGGGTTCC 3220
 Qy 1083 oLeuProleuGlyleuHlIaArPhrAlaIaIaProValleuProArpProProthrlIeSe 1103
 Db 3221 GCTTCCA-----ACAAACAGCAACCAAGGCAACCACTCCCTCATCC 3265
 Qy 1103 rAnpProProleuIleSerSerAlaYhIaProSerValleuGlunArgInIeGl 1123
 Db 3266 ATCTCTAAACACACAGTGCATCGAAGAAA--CAATCTTTATTA-----ATGGAGG 3316
 Qy 1123 YAlaIleSerGInglY-----MetSerValGlInleuHlIaValProTySe 1138
 Db 3317 GTCTATCTCAAGGAACTCTCGCACTTACTTCTTCTCATATAGAGCTTATCCACA 3376
 Qy 1138 rGlunHlIaIaIaProAlaProValIaIaProValThrMetGlyleuProleuProMetArPr 1158
 Db 3377 AGAAGCCCTTAAGCCCTCGTGGGGTCTATCTCTGGATTTGCCCGACAGAGATC 3436
 Qy 1158 oLyAlYleuAlaIaProPheSerGlyValIYsGInglInleuSerProArglYlVal1 1178
 Db 3437 TACCAAGACAGCTCTTGTACCTATACAGAGCAAGAAATTTCTCGAAGGCCAAA 3496
 Qy 1178 aglyProProgluSerleuGlyValIProthrlaGInglualaSerValleuArglYlTh 1198
 Db 3497 CTCACAACTCGAGGTTATTGGTC---AGAGCCAGCATGAAGTGTGTGACAGGAC 3553
 Qy 1198 rAlaIeGlySerValProglYlSerIleThrlYsGlyIleProSerThrArValIPr 1218
 Db 3554 TGC-----GGGGCTGTCCAAAGAGATATACCTCGGGAACTCCAGCCAGCAAAATCTC 3610
 Qy 1218 oSerArpSerAlaIeThrlYrArglYsSerIleThrlaGlyThrProAla----- 1235
 Db 3611 AGTGAAGACATTTCATCTTGGGGGCTTATTAACAGGAGACCCCACTCTGCCCA 3670
 Qy 1236 -----AspValleuTyrlYsGlyThrlIeThrArgIle---IleGlYl 1249
 Db 3671 GCGTGAATACCAACAGAGCTTTGGTGAAGGACCTGTCTCCAGATGCTTATGAAGA 3730
 Qy 1249 uArpSerProSerArglYleuArpArglYrArglYuArpSerleuProglYhIaVal11 1269
 Db 3731 AACCACTCTCGAAGAGTCT-----AGAGAGAAAGTCATCCAAAGGCGCATGTAT 3781
 Qy 1269 eTyrlGlylYlYleuGlyhIaIeIleuSerTyrlGlylYlYleuSerValThrGIncy 1289
 Db 3782 CTATGAAGCAAGTGAAGTATCTTATCATATGAATATTAATTAAGATCC----- 3833
 Qy 1289 sSerlySGlYArpSerSerGlyProProhIeGlThrAlaIaIaProlySar 1309
 Db 3834 -----CGAAGAGGAGCTCGAGTCCAAAGACGCTCATGAATAGTTTA---AAAG 3883
 Qy 1309 gThrlYrArpMetMetGlylYlYrValIaIaIaIaIeSer----- 1323
 Db 3884 AACCTATGAAGCAGTGAAGAGATATTAAGCAAGGATGTGATGAGAGTCTCTGT 3943

Qy 1324 -SerAlaSerIleGlInglYleuMetGlyArGAlaIleProProglunArghIaSerProhI 1343
 Db 3944 GTCAAGCACTTTAAAGGAGTCTGATATGCCAGCATTAACC-----AGGGAGACCTTCA 3997
 Qy 1343 v---HlIaIeUlySGlInglInhIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 1362
 Db 3998 TTCTGACCTCAAAAGAGAGAGCTGTCTGTGGTTTCATATAGCAGGACACCAAGAGC 4057
 Qy 1362 rTyrlValIa 1382
 Db 4058 CACAGCGAAAGCTTTGAGAGCGCCTT---AATATCCCAACAGATTAATAAGGAGAG 4114
 Qy 1382 YThrProProProProProSerArpArpLeuThrGluAlaIaIaIaIaIaIaIaIaIa 1402
 Db 4115 C-----CTTCCATCCAGCATTTGAAGTGCATTT----- 4145
 Qy 1402 uGlyProleuYleuYArpAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 1422
 Db 4146 -----ACCAAGAAACCA---TAGATGT---ATCAGCAGCATCAAGAAATGGG 4192
 Qy 1422 YArGSerIleHlIa 1437
 Db 4193 GCGCTCATTCATGAATCCACAGAGATATTCTAATCAGAGAAAGCCGAAACTCC 4252
 Qy 1437 oGlyleuProleuAlaIa 1457
 Db 4253 AGAAGTGTCAAGACAGAGCCCAATATTAAGGTTTCATTTCCAGGACACACCAT 4312
 Qy 1457 uLyTyArpThrlYAlaSerThrlYrArglYsSerlyYhIaIaIaIaIaIaIaIaIaIa 1477
 Db 4313 AAAATTGCAACAC---TCAGTCAATAGCATATCAACAAATGTGAAGTCTTAA 4369
 Qy 1477 eGlySerProglYArGThrPheProValHlIaIaIaIaIaIaIaIaIaIaIaIaIaIa 1496
 Db 4370 CACAGGCGCTCAGCAACTA-----CCCGTGAATCTGAAATGTACACAGAACAT 4423
 Qy 1496 aArGAlaIeGluAlaIa 1511
 Db 4424 AAAAGTATGAAGCGGAAATATATGAGTGAAGCAAGGAGGACGACATGAGGCGG 4483
 Qy 1511 gProglYThrlaIa 1531
 Db 4484 GCACAGTGAAGTGTAGCTCTGCCCCCTCGTTCAGAGTCA-----CTCA 4534
 Qy 1531 oGlyleuGlyYsProArglInSerProleuThrlYrGlunArghIaIaIaIaIaIaIa 1548
 Db 4535 CGAAGCTCCCAAGACAGCTGAGCCCGGACTATATGAGACAGCAGCCCTGCGAGAC 4594
 Qy 1549 -----ProhAlaIa 1566
 Db 4595 TCCTGTAGCTTACAGAACACCATATCCAGAGCTCCCATATAGAACGA----- 4646
 Qy 1566 rProArGleuGInglYlYsleuSerSerSerlyYs---AlaSerGlnaArpArglYle 1585
 Db 4647 -----ACTTGATGTCTTCTCCAGCAAGTGTCCAGATCATGAAGAAATC 4693
 Qy 1585 uThrSerThrlProArglu-----IleAlaYsSerProhIaIaIaIaIaIaIaIaIa 1601
 Db 4694 AACTGTGACCCCAACCAAGAAAGATATACAGAGCAAGTCTCCAGTCCCGGGGTGA 4753
 Qy 1601 oGlyhIa 1621
 Db 4754 TCCCATCGTAGTAC-----AGCCATTTGATCTCATCAGAGATAGCCGTGAGG 4807
 Qy 1621 lArpLeuTyArGSerHlIa 1641
 Db 4808 AAGAGTTATCGAGCAGCACTACCCAGCATTTGATCA---GATATGCCCTTTCAGAG 4864
 Qy 1641 eProleuArpAlaIa 1661
 Db 4865 GCGTTTGATCTCGTGTGCTTAACTGTTAAGAGACGCTTTCACAAACCCAGAGATA 4924

QY 1661 rProHisLeuTYrProProTYrLeu1LeaArgLyTYrProAspThrAla1LeuGluAs 1681
 DB 4925 CCCAAGTCAGTACCAAGCTCTAT-----GCATGGAGAA 4957
 QY 1681 n---ArgGlnThr1Le1LeaAsnAspTYr1LeThrSerGlnGlnMetHis1AsnThrAl 1700
 DB 4958 TCACAAGCGAGAAATCTCTCAAGATTAACATTACCTCACAGCAAGATGCAAGTGAATG-- 5015
 QY 1700 aThrAlaMetAlaGlnAlaAspMetLeuArgLyLeuSerProArgLySerSer1e 1720
 DB 5016 -----CGCCCTGATGTACACAGGGAGCTGTCCCAACAGAGAGCAAGCACT 5059
 QY 1720 uAlaLeuAsnTYrAlaAlaGlyProArgLy1Le1LeaAspLeuSerGlnValProHisLe 1740
 DB 5060 GGGGCTCCCTTACCCAGCT---ACAAGAGGAATCATTTGACTGACCAATATAGCTCCA-- 5114
 QY 1740 uProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTy 1760
 DB 5115 -ACAATCTTAAGTCCCTCATGACAGGGGAAAGAACACCCCTCCATGAGACAGATCACGTA 5173
 QY 1760 rLeuProThrAlaProGlnProPheSerSerArg--HisSerSerSerProLeuSerPr 1779
 DB 5174 TATTCTGTGTACAGAGTTACTTTCCCTCCAGGCCATATAAAGCTGTCTCTGTCTCC 5233
 QY 1779 oGlyLyProThrHisLeuThrLyAspProThrThrThrSerSerSerGluArgGluArgAs 1799
 DB 5234 AGGACACCCCAACACACTT-----GCAGACGCTGCAAGTGTGAGAGGGAAGAGA 5284
 QY 1799 pArgAspArgGluArgAspArgAspArgGluArgGlyuArgLySer1LeuThrSerThrTh 1819
 DB 5285 ACGGGAAAGGAGAAAGAGAGCGCAACGTAGCGCGAGCGTGAGCGTGAACGTGAACGCA 5344
 QY 1819 rThrAlaGluHisAlaPro-----1LeThrAspProGlyTYrThGluGlnSerSerG1 1836
 DB 5345 AAGGATCGCTGCTGCTCCCGCTGACCTCACTACACACAGGCTTCAGAAACAG----- 5396
 QY 1836 ySerSerGlySerSerGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHis 1856
 DB 5397 -----CCAGCGCGCCCTGGACAGCCACAGCAAGATA 5422
 QY 1856 sAlaHisGlnHisSerPro1LeSerProArgThrGlnAspAla---LeuGlnGlnArgPr 1875
 DB 5423 TGTTCGCTCCCTTCCCT---TCAGTAAAGACTCAGAGAACCATCTCTGCAACAGAGACC 5479
 QY 1875 oSerValLeuHisAsnThrGlyMetLyGly1Le1LeThrAlaValGluProSerLyAsPr 1895
 DB 5480 CAGTGTTCCTCAGAGGACCAATAGAAACCACTGATATACACACTTGGACCCAACTGCTCA 5539
 QY 1895 oThrValLeuArgSerThrSerThrSerSerProVal-----ArgProAlaAl 1911
 DB 5540 GCTACGCATCATGCCACTGCTCTGAGGGGCCCTTCCATTAAGTCAAGAGCCCTGCAAGCTTC 5599
 QY 1911 aThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrrProth 1931
 DB 5600 CCGTTAACCACTACTGCTGGAGATGCC---CTGGCTGCTCTTGTGATGCTGCAAGCTTCTGC 5656
 QY 1931 rLeuMetGluProValLeuLeuProLyGlnAlaProArg----- 1944
 DB 5657 ACCCCGATGAGATGTTCCAAAACAAAGAGACTAAGCATGAAGCTGCCAGTTAGAAGA 5716
 QY 1945 -----ValAlaArgProGluArgProArgAlaAspThrGlyHisAl 1958
 DB 5717 AAATTGAGAAAGACAGTCAAGCTAGTGAACAGCAGCAGCTAGCAAGAAAACCT 5776
 QY 1958 aPheLeuAlaTyrrProProAlaArgSerGlyLeuGluProAlaSerSerProSerLySgl 1978
 DB 5777 GAGAGTGAGAAAGATCTGTTCAAGTGCTGCTCACTTCTTCAAGCTTCAAGTGGCAA 5836
 QY 1978 ySerGluProArgProLeuVal-----ProPr 1987
 DB 5837 GGGCCAGCTCATGCTCAGTACTGATTTCTGAGGCTGGAAAGATTAAGGGCTCTCTCC 5896
 QY 1987 oValSerGlyHisAlaThr1LeAlaArgThrProAlaTyAsn-----LeuAlaProHis 2005

DB 5897 AAAATCCAGATATGAGAAAGACTAAGACCCGAGGGAAAGACTACCATTAATGCAAGCTTA 5956
 QY 2005 HisAlaSerProAspProProAlaProProAlaSerAlaSerAspProHisArgGlyTy 2025
 DB 5957 CTTCATAGACGATCATCACCCGGCAAAATTCCTCGCAACAGATGCGAGGAACGTGG 6016
 QY 2025 eThrGlnSerLyProPheSer1LeGlnGluLeuGluLeuAspArgSerLeuGlyTYrHisG1 2045
 DB 6017 CTCTCAAAAGTTCAAGCTTCT-----AGTACTGTCTTCTTCCACAG 6058
 QY 2045 ySerSerTyrrProGluGlyValGluProValSerProValSerSerProSerLeuTh 2065
 DB 6059 GTATGAAACGGGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6118
 QY 2065 rHisAspLyGlyLeuProLyHisLeuGluGluLeuAspArgLySerHisLeuGluGly1 2085
 DB 6119 CAGGAAAGACCAAGAGCTTACACCCAGACATGTTAAGGAAATCAAGAAATGA 6178
 QY 2085 uLeuArgProLyGlnProGlyProValTyLeuGlyGlyGlyAlaAlaHisLeuProHis 2105
 DB 6179 GTCCACTGCACAGATGAAGTCCA-----CTGCATCA 6211
 QY 2105 sLeuArg-----ProLeuProGluSerGlnProSerSerSerProLeuLe 2120
 DB 6212 TTATGCTCCAGCAAGAAATCAACATCTCAGAGAACAGCAACCACTGCCCATCTTC 6271
 QY 2120 uGlnThr-----AlaProGlyValLySerGlyHisGlnArgVal1ThrLeuAlaGlnHis 2138
 DB 6272 CCAATCAGAGGGAATGGGACAGCTGCCAGAGCCATGACATGATCAACATTCGTGACCA 6331
 QY 2138 sLeSerGluVal1LeThrGlnAspTYrThrArgHisHis---ProGlnGlnLeuSerAl 2157
 DB 6332 CATCTGCAAAATTAACACACAAAGATTTGCTGAATCAAGTTCCTCCAGCTTCTAC 6391
 QY 2157 aProLeu-----ProAlaProLeuTyrrSerPheProGlyAlaSerCysProVa 2173
 DB 6392 TTCTACATTCCAAACCTCAACATCTGCTTGTGATCAACACCT----- 6434
 QY 2173 lLeuAspLeuArgArgProProSerAspLeuTyrrLeuProProProAsp----- 2189
 DB 6435 -----GTAAGAACCTAAACCTCAAGCCGCTACAGCCCAAGATCAACATCTCAAGCTGT 6487
 QY 2190 ---HisGlyAlaProAla---ArgGlySerProHis-----SerGluG1 2202
 DB 6488 CTTCGATCCCAAGACAGGCTCTAGAGTCTCTCAGAAATCTTGTGATTAATTCCTGGGG 6547
 QY 2202 yGlyTyAspSerProGluProAsnLyThrSerValLeuGlyGlyGlyGluAspLyT1 2222
 DB 6548 AAGCAGGCTGGAATAATCTCCAGAGAGAGTCAATTC-----CAATCAAGAGCCCTA 6598
 QY 2222 eGluProValSerProProGluGlyMetThrGluProGlyHis1SerArgSerAlaValTy 2242
 DB 6599 TGAAGCCATCTCCCAACCCCAAGGCCCTGCTGTG-----CATGAAGACAGAGACGAT 6652
 QY 2242 rProLeuLeuTYrArgAspGlyGluGlnThrGluProSerArgMetGlySerLySerPr 2262
 DB 6653 GTTGCTCTGTGACAGAGGGAGTGAGACCTCTGACGAAGAGAGTCAATTCGATCAACC 6712
 QY 2262 oGlyAsnThrSerGlnProProAlaPhePheSerLyLeuThrGlnSerAspSerAlaMe 2282
 DB 6713 AGAAGATATTAACCTACTTCTTCAATTCACAAAGCTT---GAAAGCATCAACCCAT 6769
 QY 2282 tValLySerLyLyGlnGlu1LeAsnLyLySerLeuAsnThrHisAsnArgAsnGluPr 2302
 DB 6770 GGTAAATATTAAGAAACAGAAATTTTGTGAAGTGAAGTCTCTCTGAGGAGTGATCTC 6829
 QY 2302 oGluTyrrAsnHisSerGlnProGlyTYrThGlu1LePheAsnMetProAla1LeThrGlyTh 2322
 DB 6830 TGATATGACAGCTGCTCAGCCACAGAAACAGAGATCTTCAATCTGCGCCAGATTACCAATC 6889
 QY 2322 rGlyLeuMetThrTYrArgSerGlnAlaValGlnGlnHisAlaSerThrAspMetGlyLe 2442

Db	1291	AAACAAGAGAAACAAGCAAGAAAGATTTCAG---	CGAGTGGGCAAGAGGCAAGCTGGCTT	1347
Oy	362	SerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGlu		381
Db	1348	TCAGGCAACCATTTGGCTAGAGAGTGAAGATGATGATTTCTGAAATTAATTTGATGGGCTCTCGAG		1407
Oy	382	GInGluAenLeuGluValysGlnMetArgGlnLeuAlaValIleProPomMetLeuTyrAsp		401
Db	1408	CAGAGAGAAATTAATGAGAAACAAATAGCGGAGCTCTCTGGAATTCACCTATGATGTTGAT		1467
Oy	402	AlAspGInGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspPomMetLys		421
Db	1468	GCAGAAACAAGACGAGTCAAGTTCATTAACATGATGGGCTTAATGAGAGAACCTATGAAA		1527
Oy	442	ValTyrLysAspArgGlnValMetAsnMetIrrpSerGluGlnGluValysGluThrPheArg		441
Db	1528	GTGATTAAGAATAGGCAAGTTCATATGATATTTTGGACTGACCATGAAAGAGAACTCTTAAG		1587
Oy	442	GluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluValArgLys		461
Db	1588	GACAAAGTTATTCACGACATCCAAAACCTTTGGACTAAATTGCATCATACTTGGAGAGGAAG		1647
Oy	462	ThrValAlaGluCysValLeuTyrTyrTyrIleuThrLysLysAsnGluAsnTyrLysSer		481
Db	1648	AGTGTTCTCGATGTGTGTGTGTGATTACATTTAAACCAAGAAAATATGAAATTTAAAGCC		1707
Oy	482	LeuValAlaArgArgSerTyr---ArgArgArgGlyLysSerGlnGlnGlnGlnGln		500
Db	1708	CTCGTCAGAAAGAAATTAATGAGAAACGACAGAGCAGAAACCAAGCAATTGCTCGAACCTCG		1767
Oy	501	GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGluLys		520
Db	1768	CAAGAGAGAAAAAGTGAAGAAAAAGAAAG---GATTAAGACAGAAAACAGAAAAA		1821
Oy	521	AspGluLysGluLysGluLysGluLysGluLysGluGluGluLysProGluValGluAsn		540
Db	1822	AAAGGAAGAAAGAAAGAAAGTGAAGAGAAAAAGTGAAGAAAAAGAACTCCAAAGAAAT		1881
Oy	541	AspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGlu		560
Db	1882	ACCAAGGAA-----AAGGACACAGATAGATGTTACAGCA---GAGAAACCTGAGGAA		1929
Oy	561	LysGluValaValAlaSerLysGlyValArgLysThrAlaAsnSerGlnGlyValArgLysGly		580
Db	1930	AGAGGCAAGCCACACCCCGGGGGCGGAAGAACTGCCAACATCGGGCGCGCTTAAGGGC		1989
Oy	581	ArgIleThrArgSerMetAlaAsnGluLysAsnSerGluGluAlaIleThrProGlnGln		600
Db	1990	CGGATCACCAAGTCCATGACAAACGAAGCTGACAGTGGCAGTGGCGACCGCAGCGGCT		2049
Oy	601	Ser-----AlaGluLeuAlaSerMetGluLeuAsn		610
Db	2050	ACTGAAGAGCCCCACCACTCTGGCACCGCCACACAGAACCCATTTCACAGAGCCTGTG		2109
Oy	611	GluSerSerArgTrrpThrGluGlnGluMetGluThrAlaLysLysGlyLeuLeuGlnHis		630
Db	2110	GAGACCTCTCCATGAGCAAGAAAGAAATATGAAATGCTTAAGAAAAAGCTTGATGAACAT		2169
Oy	631	GlyArgAsnIrrpSerAlaIleAlaArgMetValGlySerLysThrValSerGluCysLys		650
Db	2170	GGTGCTAACTGGGCGACGAATTCCTAAATTTGGTGGGAGAAAGAAAGTCAATCTTAATA		2229
Oy	651	AsnPheTyrPheAsnTyrLysLysArgGluAsnLeuAspGluIleLeuGlnGlnHisLys		670
Db	2230	AACTCTCTATTTTAACATATAAAGGCGACACATCTTGAACAACCTCTTCACACAGATATAA		2289
Oy	671	LeuLysMetGluLysGluValArgAsnAlaArgArgLysLysLysValaProAlaAlaIle		690
Db	2290	CAGAAAAACTCACGAAAAACCTCTGAAGACGAGATGTGTCTCAATGTGGAAGTGTGCT		2349
Oy	691	SerGluGluValaIlePheProValValGluAspGlnGluMetGluAlaSerGlyVal		710
Db	2350	TTCCATGTTTTCGGCT-----CAGGAGGATGGAATTAATGAAGCCTCC-----		2391

[illegible]

Oy	1020	ArgGluYlySerSerArgSerProAlaProProAlaAspLysGluAlaPheAlaIaGluAla	1033
Db	3220	TGTGGCAATCCAAAGATGCCA-----AACAGAGAG-----	3249
Oy	1040	GlnLysLeuProGluYAspProProCysTrpThrSerGlyLeuProPheProValProPro	1055
Db	3250	-----TGG-----	3257
Oy	1060	ArgGluValIleLysAlaSerProHlaAlaProAspProSerAlaPheSerTyrAlaPro	1079
Db	3253	---GAAGTCCTTCACGCTGTCACAT-----CAATTGATTAACAAATCTCCT	3297
Oy	1080	ProGluYAspProLeuProLeuGlyLeuHlaAspThrAlaArgProValLeuProArgPro	1099
Db	3298	GAAGGCGTTCGGCTTCGG-----ACAACTCCGACCAACGAGGCCACCGGCC	3342
Oy	1100	ProThrIleSerAspAspProProProLeuIleSerSerAlaLysHlaProSerValLeuGlu	1119
Db	3343	CCTTCATCTCCGTCTACCCAAACCAACAGATGGCTTCGAAAAA--CAATCTTTATA--	3366
Oy	1120	ArgGlnIleGlyAlaIleSerGlnGly-----MetSerValGlnLeuHla	1134
Db	3397	---ATGGAGGGCTCATCTATCACAGGGAACACAGGCACTTATTGACTTCTCATTAATCAG	3453
Oy	1135	ValProTyrSerGluHlaIleLysAlaPro--ValGlyProValThrMetGlyLeuPro	1153
Db	3454	GCTTCCTACACTCAAGAAACACCCAAAGCCGTCAGTAGATCTATCTCTTGACTGCCA	3513
Oy	1154	LeuProMetAspProLysLysLysLeuAlaProPheSerGlyValIleGlnGlnLeuSer	1173
Db	3514	CGGCACAGGAATCTGCCAAATTCAGACTTTGGCCCTACATCAACAGAAAGAAATTTCT	3573
Oy	1174	ProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnAlaSer	1193
Db	3574	CCCCGAAGCCAAACTCAACACTGAGGGTCTGTGGTC---AGGCCCCAACATGAAGT	3630
Oy	1194	ValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIlePro	1213
Db	3631	GTAACTCAGAGGTACCGCA--GGAGCCATCAAGAAAGAAATATCACTCGGGAATCTCA	3687
Oy	1214	SerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHlaGlyThr	1233
Db	3688	ACCGAACAAATTTCAGTGAGAGCATTCATCCCTACCGGGCTCTATCTACACAGGCAAC	3747
Oy	1234	ProAla-----AspValIleTyrLysGlyThrIleThrArg	1245
Db	3748	CCGGCTCTGCCCCACAGCTCGCATACCAACAGAGCTTTGGTGAAGGGGTCCATTTCAGA	3807
Oy	1246	IleIleGlyGluAspSerProSerArgLeuAspArgIleArgGluAspSerLeuProLys	1265
Db	3808	ATGCCCATTTGAAGCAGACAGTCTT-----CAGAAAGCCAGAGAGGAATGCATCCAA	3861
Oy	1266	GlyHlaIleValIleTyrGlnGlyLysLysAlaIleLeuSerTyrGlnGlyGlyMetSer	1285
Db	3862	GGCCATGTTATTTTGAAGGCAAAAGTGACATATCTTGATATGATTAATTAAAGAT	3921
Oy	1286	ValThrGlnCysSerLysGluAspGlyYArgSerSerSerGlyProProHlaGluThrAla	1305
Db	3922	GCC-----CGAAGAGGACTTGAGAGTCCAGACAAAGCTCATGAATACGT	3966
Oy	1306	AlaProLysAspArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSer----	1323
Db	3967	TTA---AAGAGAACTATGAAATCAGTGAGAGAAATTAATTAAGCAAGGATGTCAATGAG	4023
Oy	1324	-----SerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluLys	1339
Db	4024	GAGTCTCCTGTATCAGCACCGTTGAGGGGTGATATGCCAGACATTAACCC-----AGG	4077
Oy	1340	HlaSerProHla---HlaLeuLysGluGlnHlaHlaIleAspGlySerIleThrGlnGly	1358
Db	4078	GGAGATCTCTATTCGACTCAAGAAAGAAAGCACTGATATGCTTGCTGCATATATCAGAGG	4137
Oy	1359	IleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeu	1378

Db	4138	ACACCAAGCAACCAACTGAAAGCTTTGAMATGGCCCTT---	AAATATCCCAACCAAAATT	4194
Oy	1379	LYSARGGLUGLYTHRPROPROBROPROBROSERATGAPBLEUTHRGUVALATYLYS		1398
Db	4135	AAAAAGGAAGT-----	CTCCCATACAGCATTTGAAGGTCCATT---	4233
Oy	1399	ThRGlnAlaLeuGlyProLeuLybLeuYafProAlaHisGluGlyLeuValAlaThVal		1418
Db	4237	-----	ACCAAGGAAMACCA---TATGATGGC---ATCCACCATTC	4272
Oy	1419	LYSGUVALAGLYARGSERILEHISGLULERPROARGGLUGLYLEU-----		1433
Db	4273	AAAGAAATGGGGCGCTTCATTCATGACGATTTCCAAAGGCAAGATATTTTAATCAGGAAGT		4333
Oy	1434	ARGHISTHRPROGLULEUProLeuAlaProArgProProLeuLybGluGlySerILEthGIN		1455
Db	4333	CGGAATACTCCAGAAAGTGCTCAGAGCACAGCGCGAATATTAAGGGTTCATTTCCAG		4399
Oy	1454	GLYTHRProLeuLybTYRASPThrGLYAlaSerThrThrGlySerLYbHisbAPVAL		1473
Db	4393	GGCACACCAATPAAAGTTTGACAAACAC---TCAGGTCAATCGGCATCAACACAAATGTC		4449
Oy	1474	ArgSerILEuILEGlySerPROGLYArgThrThreProProVALHisProLeuAPVALMet		1499
Db	4450	AAATCTCTTAATCACGGGGCTTAGCAAACTATCCCGTGGAATGCTCCGCTGAAATGTGTG		4509
Oy	1494	AlaAPAR---AlaArgAlaLeuGlnAlaArgAlaLeuTYRGLU-----	GLISER	1507
Db	4510	CCAGAGCAATPAAAGTGCTGAAGAACGGGGAAATATGATGAGTGTGAAGCAGCGAGACC		4568
Oy	1508	LeuLYSserATGPRGLYThralaSerSerSerGlyGlySerILEalArgGLYAlaPro		1527
Db	4570	GTGGCTTCCCGGCACACGTGAGTGTAAGCTCTGGCCCTCGCTTTAGGTCCACA---		4626
Oy	1528	ValILEVALProGluLEuGlyLYbProArgGlnSerProLeuThrTYRGLNAPHisGLY		1547
Db	4627	-----CTGCATGAAAGCTCCCAAGACACAACTGACGCTGGGATTTATGATGACACACAGT		4680
Oy	1548	Ala-----	ProPhaLEGLYHisLeuProAlaArgGlySerProVALThMet	1563
Db	4661	GCACGGAGACCCCTGTGAGTTATCAAAACACCACTGCCAGAGGTCCACCATATGAC		4740
Oy	1563	ArgGLUProThrProArgLEuGlnGLYSerLeuSerSerILEYAlaSerGlnAPAR		1582
Db	4741	AGAACTTGTGATGTTACATTCCT-----	CTTAACAAGTCAACCATCATGAA	4788
Oy	1583	ArgLYbLeuThrSerThrProArgGLU-----	ILEAlaLYbSerProHisSER	1598
Db	4789	AGGAAATCGACACTGACCCCTTACCCAGAGGAAAGTATCCACAGCGAAGTCTCCAGTGCCT		4848
Oy	1599	ThrVALProGlnHisbLeuProHisProILESerProTYRGLNHisbLeuAlaGLYVAL		1618
Db	4849	GGGGTGAACCTGTGCTGAAGCAC---AGTCCGTTTATCCCATCAACAGGGAGGC		4902
Oy	1619	SerGLYVALAPLeuTYRArgSerHisILEProLeuAlaPheAPARProThrSerILEPro		1638
Db	4903	ACTGCAGGCGAGGTTTATTTGAGGCCACTGCCCCAGCAATTGATTCGA---GCCATGCTCT		4959
Oy	1639	ArgGLYILEProLeuAPAR---AlaAlaAlaAlaTYRLeuProArgHisbLeuAlaPro		1657
Db	4960	TTTCAACAGGGCTTTGATCTCGACGGCGGTCTTAACGTTTCCAGACAGACTTTCACCA		5019
Oy	1658	AsnProThrTYRProHisbLeuTYRProProTYRILEuLEArgGLYTYRProAPARThrala		1677
Db	5020	ACTTCAGATTACCCAACTCAAGTACACCTTTAC-----		5052
Oy	1678	AlaLeuGlnAPAR---ArgGLINTHRILEAlaAPARbTYRILEThrSerGlnMetHis		1696
Db	5053	GCATGTGAGAAACAAACAGACAACTTTAAATGATTAACATTACTCAACACAGATGCAA		5112
Oy	1697	HisbANThrAlaThrAlaMetAlaGlnArgAlaAPARMetLeuArgGLYLeuSerProArg		1716

Db 5113 GTGAAGCTTG-----CGTCCAGATGTGGCCAGAGACTCTCCCCAGA 5154
QY 1717 GtuserSerLeuAlaLeuantiTyraAlaAglyProArgGlyTyleLeuSerIn 1736
Db 5155 GAGCAGCCACTGGGTCTCCCATACCCAGCA---ACGAGAGAACTTAACTGACCAAT 5211
QY 1737 ValProHisLeuProValLeuValProProthrProGlyThyProAlaThrAlaMetAap 1756
Db 5212 ATGCCTCCA---ACAATTTAAGTCTCATCCAGGGGGAACAAGACTCTCCCATGAGC 5268
QY 1757 ArgLeuAlaTyLeuProThrAlaProGlnProPhseSerArg---HisSerSer 1775
Db 5269 AGAATCACTTAATATTCCTGTACACAGATTACTTCCCTCCAGGCGGTACAACCTGTGT 5328
QY 1776 ProLeuSerProGlyGlyProThrHisLeuThrLyProThrThySerSerSerGlu 1795
Db 5329 TCCATGTCTCCAGACCCCAACACACTT-----GCACTGCTGCAAGTGTGAG 5379
QY 1796 ArgGluArgAspArgAspArgGluArgAspArgGluArgGlySerIleLeu 1815
Db 5380 AGGGAACGGGAACGGGAGCGGAGAGAGCGGAGCGGGAACG-----ATTGCT 5430
QY 1816 ThrSerThyThyValGluHisAlaProIleTrrArgProGlyThyGluGlnSerSer 1835
Db 5431 GCAGCTTCTCCGACCTC-----TACCTGCGGCGAGCTCCAGAACAG----- 5472
QY 1836 GlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgProAlaSerHisSer 1855
Db 5473 -----CTGCGCGACCTGCGACCTCAGA 5496
QY 1856 HisAlaHisGlnHisSerProIleSerProArgThyGlnAspAla---LeuGlnGlnArg 1874
Db 5497 TATGTTCCGCTCCCTCCCTC---TCAGTAAGAACTCAGAGACCATGTTGCAACAGAGA 5553
QY 1875 ProSerValLeuHisAsnThrGlyMetLyGlyIleIleThrAlaValGluProSerLy 1894
Db 5554 CCCAGGTGTTTCCAGGAACCAATGGAACCAAGTGAATCAACCTTGAATCCAACTGCT 5613
QY 1895 ProThrValLeuArgSerThrSerThrSerProVal-----ArgProAla 1910
Db 5614 CAGCTACGAATCATGCGCACTGCTGCTGGGGGCGCTTCAATAGCCAAAGCGCCCAAGCC 5673
QY 1911 AlaThrPheProProAlaThrHisCyProLeuGlyGlyThyLeuAspGlyValTyPro 1930
Db 5674 TCCCGTTACAACTGCTGGGATGCC---CTGGCTGCTGTTGTGATGCTGCAAGCTCT 5730
QY 1931 ThrLeuMetGluProValLeuLeuProLy-----GluAlaProArgValAla 1946
Db 5731 GCAACCCAGATGATGTGTCCAAACAAAAGAGATGAAGCATGAAGCTGCCAGTTAGAA 5790
QY 1947 ArgProGluArgProArgAlaAap----- 1954
Db 5791 GAAATTTAGAAAGCAGGTCAAGCAAGATTAGTGAACAGACAGACTAGACGAAGAAC 5850
QY 1955 -----ThyGlyHisAlaPheLeuAlaLy 1962
Db 5851 CTGAGAGTGAGAAAGATCTGTTCAGTGTTATACACTTCTTCAAGCTTCCAAAGTGGC 5910
QY 1963 ProProAlaArgSerGlyLeuGluProAlaSerSer-----ProserLyGlySer 1979
Db 5911 AAGCCC-----CAGCTCATTTCTTCAGTAGATTATTCTGAGGCTGGGAAA 5955
QY 1980 GluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAla 1999
Db 5956 GATTAAGGGCT-----CTTCCAAATCCAGATATGAGAGAGAGCTAAGACCAAGGG 6009
QY 2000 LyAsn-----LeuAlaProHisHisAlaSerProAspProProAlaProProAlaSer 2017
Db 6010 AAGACTACCACTTACTGCAAGTAACTTCAATGACGTATCATCCCGGCAAAATGTGCTCG 6069
QY 2018 AlaSerAspProHisArgGlyLyThrGlnSerLyProPhseSerIleGlnGluLeuGlu 2037
Db 6070 GACAAAGATGCGAGGAACTGGCTCTCAAAAGTTCAAGACTTTCT----- 6114

QY 2038 LeuArgSerLeuGlyTyHisGlySerSerTySerProGluGlyValGluProValSer 2057
Db 6115 ---AGTAGCTTATCTTCCACAGGATATGAACACCTAGAGATGATTAAGGCTGAATAGT 6171
QY 2058 ProValSerSerProSerLeuThrHisValeryGlyLeuProLyHisLeuGluGluLeu 2077
Db 6172 CTGGCAGCTCACTCGCGCACCCAGAGAACTGCAAGCACTTATCAGCCAGAGGTTGTT 6231
QY 2078 AspLySerHisLeuGluGlyGluLeuArgProLyGlnProGlyProValLyLeuGly 2097
Db 6232 AAGCAAAATCAAGCGAAATATCTTACCAACAATATGAAGACA----- 6279
QY 2098 GlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSer 2117
Db 6280 -----TTACATCACTATGACCA-----CAGCAGGAATCACCATCT 6315
QY 2118 ProLeuLeuGlnThrAlaProLy-----ValLyGlyHis 2129
Db 6316 CCCCAACAACAGCTGCGCCCTTCTTCAAGAGAGAGGAATGGGCAAGTGGCCAGAGACC 6375
QY 2130 GlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyThrArg 2149
Db 6376 CATCGCTGATCACTTCTGATCATCATCTGTCAAATATATCACAAAGATTTTGTTA 6435
QY 2150 HisHis-----ProGlnGlnLeuSerAlaProLeuProAlaProLeuTy 2164
Db 6436 AATCAAGTTTCTCGCACTCCCAAGC-----CTTCCTTCTTCACTTACATTCAG 6486
QY 2165 SerPheProGlyAla-----SerCyProValLeuAspLeuArgArgProProSerAap 2182
Db 6487 AACTCACTTCTGCTTGGTATCTACCTGTG-----AGAGCTAAACATCAAAAC 6537
QY 2183 LeuTyLeuProProProAap-----HisGlyAlaPro--AlaArgGly 2196
Db 6538 CATTACAGCCCAAGATCCAGGCTCAGTGTTCATCATCAAAAGACCAAGGTTCAAGGTC 6597
QY 2197 SerProHis-----SerGluGlyLyAspArgSerProGluProAsnLy 2211
Db 6598 TCTCCAGAAATCTGTGGACAAATCCAGGGGAAGAGCGCTGAAATCCCAAGAGAG 6657
QY 2212 ThrSerValLeuGlyGlyGlyLeuAspGlyIleGluProValSerProProGluGlyMet 2231
Db 6658 AGTCACGTC-----TCTTCCAGCCCTACGAGCCCATCTCCACCCAG----- 6702
QY 2232 ThrGluProGlyHisSerArgSerAlaValTyProLeuLeuTyArgAspGlyGluGln 2251
Db 6703 GTTCCGTTGTGATGAGAAACAGACAGACTTGCTCTTGTCTCAGAGGGCGGAGAG 6762
QY 2252 ThrGluProSerArgMetGlySerLySerProGlyAsnThrSerGlnProProAlaPhe 2271
Db 6763 CTGAGAGACAGAGAAATGATCCCGCTCACAGGAGATGAAGTAACTTGGCTTCATTC 6822
QY 2272 PheSerLyLeuThrGluSerAsnSerAlaMetValySerLyLyGlnGluLeuAsn 2291
Db 6823 TTCACCAAGCTT---GAAATACATCACCACTAGTTAAATCAAAAGAAAGAGATTTT 6879
QY 2292 LyLyLyLeuAsnThrHisAsnArgAsnGluProGluTyAsnIleSerGlnProGlyThr 2311
Db 6880 CTAAGTTAACTCTCTGTGAGAGTACCTGATGATGAGACAGCTCTCAGCAGAACT 6939
QY 2312 GluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyArgSerGlnAla 2331
Db 6940 GAGATCTTAATCTGCAAGTAACTACTGACAGGCTCAGTTAGCTCTAGAGCCATTCT 6999
QY 2332 ValGlnGluHisAlaSerThrAsnMetGlyLeuGluValIleIleArgLyAsnAlaMet 2351
Db 7000 TTGCTGATCTGCGCAGT---AATCTTGGCTGGAAGACATTATTCGAAAGCTTCATG 7056
QY 2352 GlyLyTyArgGlnTrrGluGlu-----SerProProLeuSerAla 2365
Db 7057 GAAAGCTTATGACAAAGTAAAGATCATGGAAGTTGATGATGCTTCAAGCTTATGGAGAT 7116

Db 1408 CAGGAGAAATATGAGAAACAATGCGGACGCTCTGTGATTCACCTATGATGTTGAT 1467
QY 402 AIAAspGlnGlnArgIleIysPheIleAsnMetAsnGlyLeuMetAlaAspPromeClys 421
Db 1468 GCAGAAACAAGACGATCAAGTTCACTTAACATGAAGGGCTTATGAGAGACCTATAGAA 1527
QY 422 ValIYrLysAspArgGlnValMetAsnMetTrpSerGlnGlnGluIysGluThrPheArg 441
Db 1528 GTGTATTAAGATAGGACGTTTATGATGTTTGGACTGACCATGCAAAAAGGAGATCTTAAAG 1587
QY 442 GlnIysPheMetGlnHisProIysAsnPheGlyLeuIleAlaSerPheLeuGluArgLys 461
Db 1588 GACAACTTATTCAGACATCAAAAACCTTGGACTATTTGATCATATCTTGGAGAGAAAG 1647
QY 462 ThrValAlaGluCysValLeuIYrIYrIYrLeuThrLysAsnGluAsnIYrLysSer 481
Db 1648 AGGTTCCTGATGTCTTTTGTATTTACTATTTTACCAAGAAAATGAGATTTAAAGCC 1707
QY 482 LeuValArgArgSerIYr---ArgArgArgGlyLysSerGlnGlnGlnGlnGln 500
Db 1708 CTGCTCAGAAAGAAATTATGGAACGACAGGCAAAACGCAAAATTGCTCGACCTCG 1767
QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
Db 1768 CAGAGAAAGAAAGTACAGAGAAAGAAAGAG---GATTAACAGAAAAACAGAAAAA 1821
QY 521 AspGluLysGluLysGluLysGluLysGluLysGluLysGluLysProGluValGluAsn 540
Db 1822 AAAGAAAGAAAGAAAGAAAGTGAAGAGAAAGAAAGATGAAGAAAGAACCTCCAAAGAAAT 1881
QY 541 AspLysGluAspLeuLeuLysGluLysIYrThrAspAspThrSerGlyLysPheAsnAspGlu 560
Db 1882 ACCAAGGAA---AAGGCAAGATATGATGTACAGCA---GAGAAATCTGAGGAA 1929
QY 561 LysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlnIYrArgArgLysGly 580
Db 1930 AGAGCAAAAGCAACACCCGGGGGCGAAAGACTGCAACAGTACGAGGCGCGCTTAAGAGG 1989
QY 581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluLysIleThrProGlnGln 600
Db 1990 CGGATACCAAGTCCATGACAAACGAAAGCTGCAGCTGCCAGCTCAGCCGAGGGGCT 2049
QY 601 Ser-----AlaGluLeuAlaSerMetGluLeuAsn 610
Db 2050 ACTGAAGAGCCCAACCACTCTGCCACCGCCACCAACCACTTCTACAGAGCTGTG 2109
QY 611 GluSerSerArgTrpThrGlnGlnGlnMetGluThrAlaLysLysGlyLeuLeuGlnHis 630
Db 2110 GGAACCTCTCGATGAGCAGAGAGAGAAATGGAAGTTGCTTAAGAAAGGCTTAAGAGAACAT 2169
QY 631 GlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLys 650
Db 2170 GGTCCGTAACCTGGGCAACAATTGCTAAATGCTGGGAACGAAAGTGAAGCTCAATTTAA 2229
QY 651 AsnPheIYrPheAsnIYrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLys 670
Db 2230 AACTCTATTTTAACTATTAAGCGACACAATCTTGACCAACCTCTTACAGCAGCATATA 2289
QY 671 LeuLysMetGluLysGlnArgAsnAlaArgArgLysLysLysValProAlaAlaAla 690
Db 2290 CAGAAAACTTACGAAAACTCTGTGAAGCGAGATGTCTTAAGTGAAGTGTGCT 2349
QY 691 SerGlnGluAlaAlaPheProProValValGluAspGluGlnMetGluLysSerGlyVal 710
Db 2350 TCCACGTTTCTGCT-----CAGAGAGATGAAGATTTGAAGCTCC----- 2391
QY 711 SerGlyAsnGlnGlnMetValGlnGluAlaGluAlaLeuHisAlaSerGlyAsnGlu 730
Db 2392 ---AATGAAGAAAGAAATCCAGAAAGACAGCGAA----- 2421
QY 731 ValProArgGlyLysLysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSer 750
Db 2421 ----- 2421

QY 751 IleProSerProHisThrGluAlaAlaLysAspThrGlnLysAsnGlyProLysProPro 770
Db 2422 -----GTTGAAGCTGTCAAG----- 2436
QY 771 AlaThrLeuGlyAlaAspGlyProProProGlyProProThrProProArgArgThrSer 790
Db 2437 -----CCACCGAGACAGTCTGAAAAATCTACTTCT 2469
QY 791 ArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProProPro 810
Db 2470 CCAGGAACACAGAACCTCGGTTGAGCTTGA-----CCAGCCACGGAAACT 2517
QY 811 AlaPro---ProSerProSer---AlaProProProValValProLysGlnGluLysGlu 828
Db 2518 GCACCCAGTACATCTCCCTCTTACAGATTCCAAAGTACAAACAGCTGAAGATGAAGT 2577
QY 829 GlnGluThrAlaAlaAlaProProVal----- 837
Db 2578 GTGAGAGCCAGGTGAATGACAGCATCACTGCTGAGACAGCAGAGACATGATGAT 2637
QY 838 -----GlnGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 851
Db 2638 CAGCAGAGACAGTGTGAGAGAGGTTCTGTTGATGATCCCAACCGCTACCAAA--- 2694
QY 852 AlaValAspThrGlyLysAlaGluGluProValLysSerGlyCysThrGluGluAlaGlu 871
Db 2695 ---GCTGACTCTGTGACCTGTAAGTGAAGGTCAGAAACCATGATCTAAAGTTGAA 2751
QY 872 GlnGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeu 891
Db 2752 GGTGATATATCAAGAAAGAAAGCTTGATAGACC----- 2787
QY 892 LysAlaGluLysLysGlnGlyLysSerGlyArgAlaThrAlaLysSerSerGlyAla 911
Db 2788 ---AGTGAAGAGTGAACCTAGATGAAGATTGGTGTATGCTCAGCAAAATTAATGCC 2844
QY 912 -----ProGln---AspSerAspSerSerAlaThrCysSerAlaAspGluVal 926
Db 2845 CAAAGGCCGAGCCCACTGACAGCAACATGATTCAGTGCACGCTGACGCTGATGAG--- 2901
QY 927 AspGluAlaGlnGlyLysLysAsnArgLeuLeuSerPro-----ArgProSer 943
Db 2902 ---GATGTGATGAGAGCCAGAGAGGCGAGAAAGTTTCTATGAGACTCAAAAGCTTGA 2958
QY 944 LeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLys-----ProLeu 961
Db 2959 CTGTTAAACCCCACTGGATCTATCTGCTC---TCATCTCGTTTAAACCAATTCACATG 3015
QY 962 AspLeuLysGlnLysGlnArgAlaAlaAlaIleProProIle----- 976
Db 3016 GATCTGCCACAGACTTCAGCATGAGCTGCTGTTATCCCAATGATGTATCTGCACCCCA 3075
QY 977 -----GlnValThr 979
Db 3076 TGTAAACATCAATTGGAACCCCAAGTGAACGGCTTACTCTTACAGGACACATTTAA 3135
QY 980 LysValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaPro 999
Db 3136 GCAATGATGATGACGACTCTCTGAG----- 3162
QY 1000 ProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerPro 1019
Db 3163 ---GAGCAGCGGACAGAGAAAGAAAGATGTTGGATGTAAGAAATCTCAAGTCCA 3219
QY 1020 ArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAla 1039
Db 3220 TGTGGCATCTCAAGAGTCA-----AACAGAGAC----- 3249
QY 1040 GlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValProPro 1059
Db 3250 -----TGG----- 3252

QY 1060 ArgGluValIleLeuValAspProHisAlaProAspProSerAlaPheSerTyAlaPro 1079
 Db 3253 ---GAAGTCTTTCAGCTCCACAT-----CAATGATACATAATCTCCT 3297
 QY 1080 ProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgPro 1099
 Db 3298 GAAGGCGTTGGCTTCG-----ACMCTCGAACACCGGCCACCGCC 3342
 QY 1100 ProThrIleSerAspProProProLeuIleSerSerAlaLeuHisPheProSerValLeuGlu 1119
 Db 3343 CCTCTATCCCGTATCCAAACACACAGTGGCTTCAGAAAA---CACTCTTTATA--- 3396
 QY 1120 ArgGlnIleGlyAlaIleSerGlnGly-----MetSerValGlnLeuHis 1134
 Db 3397 ---ATGGAGGCGCTCATCTCAGAGGAACACAGGCACTTATTGTCCTCATATCAG 3453
 QY 1135 ValProTySerGlnHisValAlaValaPro---ValGlyProValIleThrMetGlyLeuPro 1153
 Db 3454 GCTTCTTACACTCAAGAAACCCACAGCCGTCAGTATCTATCTCTTGGACTGCCA 3513
 QY 1154 LeuProMetAspProLybLybLeuAlaProPheSerGlyValLybGlnGlnLeuSer 1173
 Db 3514 CGGCAACAGAAATGTCGCAATTCAGTACTTGGCCCTACATCAACAGAAATTTTCT 3573
 QY 1174 ProArgGlyGlnAlaGlyProProGlySerLeuGlyValProThrAlaGlnGlnAlaSer 1193
 Db 3574 CCCGGAAGCCAAATCACAACCTGAGGGTCTGTGGTC---AGGCGCCAAACATGAAGT 3630
 QY 1194 ValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLybGlyIlePro 1213
 Db 3631 GTAGTCAGAGTACCGCA---GGAGCCATACAGAGAGATATACCTCGGGGAATCCCA 3687
 QY 1214 SerThrArgValProSerAspSerAlaIleThrTyAlaArgGlySerIleThrHisGlyThr 1233
 Db 3688 ACCAGCAAAATTTAGTAGAGAGATTCATCCTACGCGGCTTATCATCAGAGGCAAC 3747
 QY 1234 ProAla-----AspValLeuTyLybGlyThrIleThrArg 1245
 Db 3748 CCGGCTTGGCCCGACTGGCATCCAAACAGGCTTTGGTGAAGGGGTCCATTTCAGA 3807
 QY 1246 IleIleGlyGlnAspSerProSerArgLeuAspArgGlyArgGlnAspSerLeuProLyb 1265
 Db 3808 ATGGCCATTGAAGCAGACAGTCTCT-----GAGAAAGCGAGAGAGAGTGCATCCAAA 3861
 QY 1266 GlnHisValIleTyGlnGlnLybLybGlyHisIleValLeuSerTyGlnGlnGlyMetSer 1285
 Db 3862 GGCCATGTTATTATGAAGGCAAAAGTGACATATCTTGCATATGTAATATTAAAGAT 3921
 QY 1286 ValThrGlnCysSerLybGlnAspGlyArgSerSerSerGlyProProHisGlnThrAla 1305
 Db 3922 GCC-----CGAGAAAGGAGCTAGAGCTCCAAAGAACGCTCATGAATCAGT 3966
 QY 1306 AlaProLybArgThrTyAspMetMetGlnGlyArgValGlyAlaIleSer----- 1323
 Db 3967 TTA---AAGGAAGCTATGATCAGTGAAGGAATATAAAGCAAGGATGTCATGAGG 4023
 QY 1324 -----SerAlaSerIleGlnGlyLeuMetGlyArgAlaIleProProGlnArg 1339
 Db 4024 GAGTCTCTGATACGACCGTTAGAGGGCTGTATGCGAGCATTAACCC---AGG 4077
 QY 1340 HisSerProHis---HisLeuLybGlnGlnHisIleHisLeuArgGlySerIleThrGlnGly 1358
 Db 4078 GGGAGTCTCATCTCTGACCTCAAGAAAGAGCTGTATGCTGGCTCCATATATGSCAGGG 4137
 QY 1359 IleProArgSerTyValGlnAlaGlnGlnAspTyLeuArgArgGlnAlaLybLeuLeu 1378
 Db 4138 ACACCAAGACCAACACTGAAGCTTTGAAGATGGCTT---AAATATCCCAACCAATT 4194
 QY 1379 LybArgGlnGlyThrProProProProProSerArgAspLeuThrGlnAlaTyLyb 1398
 Db 4195 AAAAGGAAAGT-----CTTCCCATACGAGCATTTGAAGTGCCATT--- 4236
 QY 1399 ThrGlnAlaLeuGlyProLeuLybLeuLybProAlaHisGlnGlyLeuValAlaThrVal 1418

Db 4237 -----ACCAAGAAACCA---TATGATGGC---ATCACCACCATC 4272
 QY 1419 LybGlnAlaGlyArgSerIleHisGlnIleProArgGlnGlnLeu----- 1433
 Db 4273 AAAGAAATGGGCGCTTCAATTCATGATTCAGAAATTCAGAAATTTTAATCAGAAAGT 4332
 QY 1434 ArgHisThrProGlnLeuProLeuAlaProArgProLeuLybGlnGlySerIleThrGln 1453
 Db 4333 CGGAAATCTCAAGAGTGCACAGACACAGCGCCGATATATGAGGTTCCATTTCAG 4392
 QY 1454 GlyThrProLeuLybTyAspThrGlnAlaSerThrThrGlySerTybHisAspVal 1473
 Db 4393 GGCACACCAATAAAGTTTGACAACAC---TCAGTCAATCTGCCATCAACCAATGTC 4449
 QY 1474 ArgSerLeuIleGlySerProGlyArgGlnPheProProValHisPheLeuAspValMet 1493
 Db 4450 AAATCTTAATCAAGGGGCTTACCAACATATCCCGTGAATGCTCCGTGAATTTGTG 4509
 QY 1494 AlaAsp---AlaArgAlaLeuGlnArgAlaCysTyGln-----GlySer 1507
 Db 4510 CAGAGAACATAAAGTGTGAACGGGGAATATGAGATGTGAAGACAGCGAGAC 4569
 QY 1508 LeuLybSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaPro 1527
 Db 4570 GTGGTTCGCGGACACGTCAGTGTAACTGTGGCCCTCGCTTTTATGATCCACA--- 4626
 QY 1528 ValIleValProGlnLeuGlyLybProArgGlnSerProLeuThrTyGlnAspHisGly 1547
 Db 4627 ---CTGCATGAAGCTCCCAAGCAACATGAGCCCTGGATTTATGATGACACAGT 4680
 QY 1548 Ala-----ProPheAlaGlnHisLeuProArgGlySerProValIleThrMet 1562
 Db 4681 GCACGAGAGACCCCTGTGAATTCATCAAAACACATGTCAGAGGCTCACCATGATGAC 4740
 QY 1563 ArgGlnProThrProArgLeuGlnGlnGlySerLeuSerSerSerLybAlaSerGlnAsp 1582
 Db 4741 AGAATCTGTAGTTCAATCTCT-----CTTAACAAAGTCAATCAATGAA 4788
 QY 1583 ArgLybLeuThrSerThrProArgGln-----IleAlaLybSerProHisSer 1598
 Db 4789 AGGAAATCGACATGACCCCTTACCAGAGGAAGATCCACAGAAATGTCAGTCC 4848
 QY 1599 ThrValProGlnHisAspProHisPheProIleSerProTyGlnHisLeuAspGlyVal 1618
 Db 4849 GGGGTGACCCCTGTGAGCCAC-----AGTCCGTTTATCTCCATCACAAGGACG 4902
 QY 1619 SerGlyValAspLeuTyArgSerHisIleProLeuAlaPheAspProThrSerIlePro 1638
 Db 4903 ACTGCAAGCGAGGTTATTGAGACCACTGCCACCAATTTGATCCA---GCCATGCT 4959
 QY 1639 ArgGlyIleProLeuAsp---AlaAlaAlaAlaTyTyTyLeuProArgHisIleLeuAlaPro 1657
 Db 4960 TTTCACAGGGCTTTGATCTCTGACGGGCTTACCTGTTTCAAGACAGCTTTTACCA 5019
 QY 1658 AspProThrTyArgProHisLeuTyTyProProTyLeuIleArgGlyTyProAspThrAla 1677
 Db 5020 ACTCCAGGTTACCCAACTCAGTATCAGCTTTAC----- 5052
 QY 1678 AlaLeuGlnAsn---ArgGlnThrIleIleAsnAspTyIleThrSerGlnGlnMetHis 1696
 Db 5053 GCAATGAGAAACACAGACACAGCAATCTTAATGATTAATCACTTCACAAACAGATGCA 5112
 QY 1697 HisAspThrAlaThrAlaMetAlaGlnAlaArgAspMetLeuArgGlyLeuSerProArg 1716
 Db 5113 GTGAACTTG-----CGTCAGATGTGGCCAGAGACTCTCCCAAG 5154
 QY 1717 GlySerSerLeuAlaLeuAsnTyAlaAlaGlyProArgGlyIleIleAspLeuSerGln 1736
 Db 5155 GAGCAGACCACTGGGTCTCCATACCAACA---ACAGAGAAATCTTATGACCTGACCAAT 5211
 QY 1737 ValProHisLeuProValLeuValProProThrProGlyThrProAlaThrAlaMetAsp 1756

Dh 5212 ATGCTCCA---ACAATTTAGTCTCATCCAGGGGGAACAAGACCTCTCCCATGAC 5268
Qy 1757 ArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArg---HisSerSer 1775
Db 5269 ABAATCAGCTTATATCTCTGGTACAGAGATTACTTTCCTCCAGCGCGGACAACTGTCT 5328
Qy 1776 ProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrHisSerSerGlu 1795
Db 5329 TCCATGTCTCCAGGACACCAACACACTT-----GCAGCTGCTCAAGTGTCTGAG 5379
Qy 1796 ArgGluArgAspArgAspArgGluArgAspArgGluArgGluLysSerIleLeu 1815
Db 5380 AAGGAACGGGAACGGAGCGGAGAGCGGAGCGGAGACG-----ATTGCT 5430
Qy 1816 ThrSerThrThrValGlnHisAlaProIleTyrArgProGlyThrGlnInSerSer 1835
Db 5431 GGAGCTTCTCCGACCTC-----TACCTGCGGCGGAGCTCAGAACG----- 5472
Qy 1836 GlySerSerGlySerSerGlyGlyGlyGlySerSerArgProAlaSerHisSer 1855
Db 5473 -----CTGCGCGGACCTGCGAGTCAATGGA 5496
Qy 1856 HisAlaHisGlnHisSerProIleSerProArgThrGlnAspAla---LeuGlnGlnArg 1874
Db 5497 TATGTTCCGCTCCCTTCCCT---TCAGTAAGAACTCAGAGACCACTGTTGCAACAGAA 5553
Qy 1875 ProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLys 1894
Db 5554 CCCAGGTTTCCCAAGGAACCAATGGAACAGTGAATCACAACCTTGGATCCAATGCT 5613
Qy 1895 ProThrValLeuArgSerThrSerThrSerSerProAla-----ArgProAla 1910
Db 5614 CAGCTACGATCATGCGCACTGCTGCGGGGCCCTTCAATAGCCAGAGCCCTGCGAGCC 5673
Qy 1911 AlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrPro 1930
Db 5674 TCCCGTTACAACAGCTGCTCGGATGCC---CTGGCTGCTTGTGGATCTGAGCTTCT 5730
Qy 1931 ThrLeuMetGluProValLeuLeuProLys-----GluAlaProArgValAla 1946
Db 5731 GCAACCCAGATGGATGTGCCAAACAAAGAGATGAAGATGAGCTGCGAGGTTAGAA 5790
Qy 1947 ArgProGluArgProArgAlaAsp----- 1954
Db 5791 GAAATTTGAGAAGACGTGCAGCAGCAGCATTAAGTGAACAGCAGCAGCTAGACGAAACC 5850
Qy 1955 -----ThrGlyHisAlaPheLeuAlaLys 1962
Db 5851 CTGGAGGTGAGAAAGATCTGTCAGTGTATTATACACTTCTTCAAGCTTTCCAACTGGC 5910
Qy 1963 ProProAlaArgSerGlyLeuGluProAlaSerSer-----ProSerLysGlySer 1979
Db 5911 AAGCCC-----CAGCTCATTTCTTCAAGTATTCTGAGCTGGGAAA 5955
Qy 1980 GluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAla 1999
Db 5956 GATTAAGGGGCT-----CCTCCAAATCCAGATATGAGGAGAGAGCTAAGACAGAGGG 6009
Qy 2000 LysAsn-----LeuAlaProHisHisAlaSerProAspProProAlaProProAlaSer 2017
Db 6010 AAGACTACCATTAATCTCAGCTAACTTCATAGACAGTATCATCCCGGCAAAATGGCTCG 6069
Qy 2018 AlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGln 2037
Db 6070 GACAAAGATCGAGGAGACGTGCTCAAAAGTTCAAGACTTCT----- 6114
Qy 2038 LeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSer 2057
Db 6115 ---AGTAGCTTAATCTTCTCAGAGTATGAACCACTAAGCATCTTGTGAGTAAAGT 6171
Qy 2058 ProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGlnGluLeu 2077
Db 6172 CTGCGAGCTCACTCTCGGACCCAGAGAAACTGCAAGACTATACAGCCAGAGGTTGTT 6231

Qy 2078 AspLysSerHisLeuGlnGlyGlyLeuArgProLysGlnProGlyProValLysLeuGly 2097
Db 6232 AAGCAAAATCAAGCCGGAATATGATCTTACACAGACAAATATGAAGACCA----- 6279
Qy 2098 GlnGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSer 2117
Db 6280 -----TTCATCATCTATGACCA-----CAGCAGGAATCACCATT 6315
Qy 2118 ProLeuLeuGlnThrAlaProGly-----ValLysGlyHis 2129
Db 6316 CCCCAACAAACAGCTGCCCCCTTCTCACAGCAGAGGGAATGGGCAAGTGCCAGACCC 6375
Qy 2130 GlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArg 2149
Db 6376 CATCGCTGATCACACTTGATCATCTGTCAAAATTTATACACAAGATTTCTCTGA 6435
Qy 2150 HisHis-----ProGlnGlnLeuSerAlaProLeuProAlaProLeuTyr 2164
Db 6436 AATCAAGTTTCTCCGACACTCCAGACG-----CTTCTTACTTCTACATTCCAG 6486
Qy 2165 SerPheProGlyAla-----SerCysProValLeuAspLeuArgArgProProSerAsp 2182
Db 6487 AACTCAGCTTGTGCTTGGTATCTACACTGT-----AGGACTAAACATCAAC 6537
Qy 2183 LeuTyrLeuProProProAsp-----HisGlyAlaPro---AlaArgGly 2196
Db 6538 CGTTACAGCCCAAAATCCAGGCTCAGTGTCTCATCTCAAAAGACAGGTTCAAGGCTC 6597
Qy 2197 SerProHis-----SerGlnGlyGlyLysArgSerProGluProAsnLys 2211
Db 6598 TCTCCAGAAATCTTGTGACAAATCCAGGGAAGTACGCCCTGGAATAATCCCAAGAGAG 6657
Qy 2212 ThrSerValLeuGlyGlyGlyLysAspGlyIleGluProValSerProProGluGlyMet 2231
Db 6658 AGTCACGTC-----TCTTCCAGCCCTACGAGCCCATCTCCCAACCCAG----- 6702
Qy 2232 ThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGln 2251
Db 6703 GTTCCGAGTTGTCATGAGAAACAGACAGCTTGTGCTTCTTTCAGAGGGGCGCAGAG 6762
Qy 2252 ThrGluProSerArgMetGlySerLysSerProGlyLysThrSerGlnProProAlaPhe 2271
Db 6763 CTGCAAGACAGAGAAATGATGCCCGCTCACAGGAGATGAAGTACTTGCCTTCAATTC 6822
Qy 2272 PheSerLysLeuThrThrGluSerAsnSerAlaMetValLysSerLysLysGlnLysLeuAsn 2291
Db 6823 TTCACCAAGCTT---GAAATATCATCACCCATGTGTTAATCAAGAAAGCAGAGATTTT 6879
Qy 2292 LysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThr 2311
Db 6880 CTTAAGTTGMACTCTCTGAGAGGTGACTGTGATATGGCAGCTCAGCAGGAACT 6939
Qy 2312 GlnIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAla 2331
Db 6940 GAGATCTTTAATCTGCAACAGATTACTACAGGCTCAGTGAAGCTCTAGAGCCATTCT 6999
Qy 2333 ValGlnGlnHisAlaSerThrAsnMetGlyLeuGlnAlaIleIleArgLysAlaLeuMet 2351
Db 7000 TTGCGTATCTGCGCAGT---AATCTGGGCTGGAAGACATTATCAGGAAGGCTCTCAG 7056
Qy 2352 GlyLysTyrAspGlnTyrGluGlu-----SerProProLeuSerAla 2365
Db 7057 GGAAGCTTTGATGACAAAGTATGAGATCATGAGTTGTCATGTCCAGCTTATGGAGATA 7116
Qy 2366 AsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAla 2385
Db 7117 -----GTGCTGTGATCTGCAACACTCAGTT-----GTGACCACT 7152
Qy 2386 AlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGly-----LysAla 2403
Db 7153 GGTGAGACACAGAAAGAGAGAGGAGCCATCATCCTCATTCAGAGAGAGTTTGCAAAACA 7212

QY 2404 LysValSerGlyArgProSerSerArgLysAlaLysSerProAlaPro-----GlyLeu 2421
 Db 7213 AAGCTGATCAGCAAGTCAAAAGACGAGAAATCTAAGTCTCTAATCTTGGGCAAGGCTAC 7272
 QY 2422 AlaSerGlyAspArgProSerValSerSerAlaHisSerGluGlyAspArg 2441
 Db 7273 TTAGGAACGAAACGGCTCTTCCAGTCTCTCTGTAATTCAGAAAGGATTCACAAAGG 7332
 QY 2442 ArgThrProLeuThrAsnArgValITrgLysAspArgProSerSerAlaGlySerThrPro 2461
 Db 7333 CAGACGCCA-----GGGTGGGCTGGGAAAGACAGGCCCTTTCAACAGGCTCAACTCAG 7386
 QY 2462 PheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyAlaMetAlaSerProPro 2481
 Db 7387 TTTCCTTATAACCTCTGACTGATGCGGATG-----CTCAGACGAGACTCCCAACA 7437
 QY 2482 ProGlyLeuProAlaGlySerGly---ProLeuAlaGlyProHis-----HisAla 2497
 Db 7438 CCGATTGCATGTGCTCTCCCTGCGGTGAACCAACACACTCTCTCAACAAACAGATC 7497
 QY 2498 TTrpAspGluGlyProLysProLeuLeuGlySerGlnTyrGluThrLeuSerAspSerGlu 2517
 Db 7498 TGGAGCGAGAGAGCTGCCCACTGCTCTCAGACAGATGACAGACCTGTGCGATGATGAT 7557
 RESULT 14
 AF044209 7940 bp mRNA linear PRI 01-SEP-1998
 LOCUS Homo sapiens nuclear receptor co-repressor N-COR mRNA, complete
 DEFINITION c88.
 ACCESSION AF044209
 VERSION AF044209.1 GI:3510602
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens (human)
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 7940)
 Wang, J., Hoehling, T., Redner, R.L., Kajigaya, S. and Liu, J.M.
 ETO, fusion partner in t(8;21) acute myeloid leukemia, represses
 transcription by interaction with the human N-Cor/msin3/Hdaci
 complex.
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (18), 10860-10865 (1998)
 MEDLINE 98393736
 PUBMED 9724795
 AUTHORS 2 (bases 1 to 7940)
 TITLE Wang, J. and Liu, J.M.
 REFERENCE Direct Submission
 JOURNAL Submitted (22-JAN-1998) Hematology Branch, National Heart, Lung and
 Bethesda, MD 20892, USA
 FEATURES
 source location/Qualifiers
 1..7940
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 241..7563
 /codon_start=1
 /product="nuclear receptor co-repressor N-Cor"
 /protein_id="AAC33550.1"
 /translation="MSSSGYPPNCGAFSTQSRYPHPSVOYTPENRHOEFAPVDPYR
 SHLEVSQASQLQOOQOOLRRPRLSLFPHGSDPDRRTSYRERHGGPVPVND
 SLESKRPLRLEQSDSHFORVSAVLPVHPLFEGRLASADAKDPAFGHGHAASSPI
 SCGPCDDNNAASKSKELEIOMDRVDEIAIKVBOILKLLKKQOOLEEAKRPE
 PEKVPSPPEVQHRSTVOITIDENRKAEAKIFEGSLKVELPLVYNPSDTKYH
 ENIKTVQVMRKLIIFPKRRNHAKRQEQIKQRYDOLMAMWAKVDLNNRRKAK
 ESTKREYKQPEIRKQBOQRFQVNGRAGLSATTAARSHEISEIIDGSEBN
 NEKQMOQLSVIPMMEDAQORVYFIMNGIMEDPMKYTRQDFPMVWTIDHEKIFPD
 KFIQHPNFGELIASYLERKSPDVCVLYVLTKNENKALVRNRYGRRNGOQIAR
 SOEKVEKEDEKAETKEKEKEKEDEKEDEKENDKEDKIDGTAEETEBEBO
 ATRGRKKTANSQGRKRKTRITRSTNTNEAASAAATAATEPPPLPPPEPSTEEVE
 TSWTEEMEVAKGLVEHGRNNAALAKWGTSEACQKVFYNYKRRHLDNLQOH
 KQTSRKPRBERDVSOCEVASTVSAQEDEDIASNEENPDESEVAIVPSBDSBN

ORIGIN
 Alignment Scores:
 Pred. No.: 1,07e-75 Length: 7940
 Score: 4147.50 Matches: 1099
 Percent Similarity: 53.33% Conservative: 341
 Best Local Similarity: 40.70% Mismatches: 785
 Query Match: 31.38% Indels: 475
 DB: Gaps: 108
 US-09-522-753-5 (1-2517) x AF044209 (1-7940)
 QY 16 GlnProArgTyrProProHisSerLeuSerTyrProValGlnIleAlaArgThr 35
 Db 289 CAAAGTCGATATCTCTCTCACTGTCAGATATTAATCCCAACCCGCCACAGCAG 348
 QY 36 AspValGlyLeuLeuGluTyrGln-----HisHisSerArgAspTyrAlaSerHisLeu 53
 Db 349 GAGTTGCGAGTCCCTGATTAATCTCTCTCACTGATGAGTGAAGTGAAGCAGCATCAGCTT 408
 QY 54 SerProGlySerIleLeuGlnProGlnArgArgArgProSerLeuSerGluPheGln 73
 Db 409 TTGCAGCAACAGCAGACGAGCAAGCTTCGAAGGCGACCTTCCTGCTTACGAATTCAC 468
 QY 74 ProGlyAsnGlnArgSerGlnGluLeuHisLeuArgProGluSerHisSerTyrLeuPro 93
 Db 469 CCAAGTCTGACAGCGCTCAAGAA-----AGAGAGAACTAGATTAGAACCGCTTATCCA 522
 QY 94 GlnLeuGlyLysSerGluMetGluPheIleGluSerTyrArgProAlaGluLeuLeu 113
 Db 523 GGCCCATCCCAAGTGAATCATGATTCAGATGAGAGGACACAGCTCTGGAACAGGT 582
 QY 114 ProAsp-----ProLeuAspArgProSerPro 122
 Db 583 TCTGATTCATCTTTCAGCGTGTCAAGTCTGCGGCTTTTGCCTTTAGGCAACCGCTGCCA 642
 QY 123 LeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThr 142
 Db 643 -----GAAGGCTGAGGGCT---TTCGACAGATGCTAAGAAAGATCAGCATTCGGA 690
 QY 143 GlyLysLeuGlu---ProValSerProProSerProProHisHisThrAspProGluLeuGlu 161
 Db 691 GGCAAACTGAAGCTCATCTCTCCAAATTTCCGGGGCAACCATGTGAGATGATCAAAAT 750

ATSRGTEPAVELEPTTETAPOSTSLAVPSTKPADESEVETOVNDISIAETAQMDV
 DQOEHAEBSGSDPPPPATKADSDVDEVRPENNHAIVBEDNTERKRLDLSAEKVER
 DEBLVAQOINACDRPEPPDNDSDSATDEADVEDGERORMPMPMSKPLMLPTGSI
 LVSSPLKPNPLDLPOLQHRAAVIPMWVSCTPCNIPITGTPVGYALVORHAKHESL
 LEBOROROBIDIECSSSTSPCTSSKSPRENMVLDPAHQILTNLDEGRVLPTRPT
 RPPPLIPSKITVASERFLMGASISQCTPGTYLISHQASTQDETPEKPSYSL
 GLPROQESASATLPYIKOEFSPRQNSQPEBLVRAQHEGVAGTAIIOGSIIR
 GTPTSKISVESIPSLRGSITOCPTALPQGIPEALAVKGSISMPLEDSPERGREBA
 ASKGVHYEGKSGHILSYNDIKVARSGTSPRAHEISLRGSYSEGANIKQKMSME
 SPVSALEGLICRLPRGSPHSDPKERTVSSGIMGTPTATTESPEDGAKYKOIR
 ESPPIAEPBATITKGRPDGITTIXEMGSIHIEPRODITQSRKTRPVQSTRIPI
 BGSISQCTPIKFDNNSQSHIKINVSLLTGEKLSRSGEPIYDDTSARTPVSYONTR
 DVAAGTISRHTSVSVSSGSPSVLTLHEAPKQSLGAYDDTSARTPVSYONTR
 GSPMNRSTSVITPPMNSTHNEKSTLTPTQRAISIPAKSPVGVDPVSVSPDPHNR
 GSPAGVYMSHLPTOLDPAMPFRALDPAALVAFOROLSPTEGVSPOLYMYEMTR
 OTLINDITNSQOMQVNLPRDVARGSLPREQPLGPPYPRAGIIDLTMPTITLVPHG
 GISTPPMDRITTYETQITTPPPPVNSAGSPHPTHLAAASAEERERERERER
 ERIIAASSDLYLTPGSEOPGRPGSHGVYSPSVVTOETMLQQRSPVFGTGTSTVI
 TPLDPTAOALIMPLPAGGSPISQGLPASHRYNTADALALVDAASAAPMDVAKTES
 KHEARLEENLRSRSAVSEOOQLEQKTEVERGSVOCUTYSAPFSGKQPPHSVY
 SEAGKQKGPPEKSRYEELRTGKTTITANFIDVITROIASDKOAREGSSQSDS
 SLSSHYETPSDALEVISPASPARBOETQYQYQYQYQYQYQYQYQYQYQYQYQY
 YRQOQSPSTQOOLPPSSQABEGQYPTRLTLTLADHICQITTDOPARQVSSQTPQ
 QPRTSTFQNSPSALVSTPVRTKTSNRYSPESQMSVHQRGSRVSPENLVKSRGR
 PKSPERSHVSSEPEPIIPPQYVVEHKDSSLLOSRAEPAEORNDARSQSGISY
 LPSEFTKLENTSPMVKSKKQELFRKLNSGSGSDMAAOPTEIFNLPAVTVSGVS
 SRGHSFADPASNGLIEDIRKALMGSEFDVEHGVMSOPMVEGTANTSVTSGE
 TREEDBSPHSGGVCKPKLISKNSRKSPIPGGYVIGTERPSVSVSHSGDTHR
 QTRGAMWEDRPSSTGTOPPYNPLTRMLSTSTPTPIACAPSAVNQAAHPQNRIMER
 EPAPLLSAQYETLSDSD"

```

OY      162 LeuValProProArgLeuSerIysGluLeuIleGlnAsnMetAspArgValAspArg 181
DB      751 GGTTCACCTTCAAAATCTCAAAAGAGATTAAATCAGAGATGATGCTGTAGATCGA 810
OY      182 GluIleThrMetValGluGlnGlnIleSerIysLeuIysIleValGluGlnIleGlu 201
DB      811 GAAATTTGCAAAAGTAAAGCAGATCTTAAACTGAAAGAAACAAACAGCTGTGAA 870
OY      202 GluGluAlaAlaLysProGluProGluLysProValSerProProIleGluSer 221
DB      871 GAAAGGCGAGCTTAACCTCTGAGCTGAGAGACCCGTGCTCCCTCTCTCTGAGCGCAG 930
OY      222 LysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgIysIysAlaGluAla 241
DB      931 AAACACCGCAGATATTGTCCAAATTATTATGATGACATCCGAAAACAGAAAGAAAGCT 990
OY      242 HisArgIleLeuGluGluIleGluIleProGlnValGluLeuProLeuTyrAsnGlnProSer 261
DB      991 CATAAATTTTGAAGGCTGTGGCCCAAAAGTTGAACCTGCATATTAACCGCATCA 1050
OY      262 AspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgIysIysLeuIle 281
DB      1051 GATACCAAGGTGACCATGAGAACATCAAGACAAACAGGTGATGAGAGAAAATCATATT 1110
OY      282 LeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGlnArg 301
DB      1111 TTATTTTAAAGAAAGAAATCATGCAAGAAAACAAAGGAAACAAATAATCTGCCAGCT 1170
OY      302 TyrAspGlnLeuMetGluAlaLeuGluIysLysValGluArgIleGluAsnAsnProArg 321
DB      1171 TATGATCACCTCATGAGGCAATGGGAGAAAAGTGAGACAGATTAATAATATCTCTCG 1230
OY      322 ArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArg 341
DB      1231 AGGAAAGGCTTAAGAAAGCAAAACAGAGAAATCTATGAAAGCAAGTTTCCAGAAATTCAG 1290
OY      342 LysGlnArgGluLeuGlnIleLysArgMetGlnSerArgValGlyGlnArgIysSerGlyLeu 361
DB      1291 AAACAAAGAGAACAGACAAAGATTTCAG---CGAGTTGGCGCAGAGGGAGCTGTCTTT 1347
OY      362 SerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGlu 381
DB      1348 TCAGCCACCATGCTCTGAGGTGAGCATGAGATTTCGAATAATTATGAGGCTCTCTGAG 1407
OY      382 GlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAsp 401
DB      1408 CAGGAGAAATTAATGAGAAACAAATGCGCGAGCTCTGTGATTCACCTATGATGTTGAT 1467
OY      402 AlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLys 421
DB      1468 GCAGAAACAAAGACGATCAAGTTCATTAAATGAGGCTTATGAGGAGCCCTATGAGAA 1527
OY      422 ValTyrLysAspArgGlnValMetAsnMetTyrSerGluGlnGluIysGluIleThrPheArg 441
DB      1528 GTGTATAAAGATATGCGCATTTATGAATGTTTGAATGCATGACCATGAAAGAGATCTTTAAG 1587
OY      442 GlnLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGlnIleArgLys 461
DB      1588 GACAACTTTATTCAGCATCCAAAAACCTTTGACATAATTCATATCTTGAGAGAGAG 1647
OY      462 ThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSer 481
DB      1648 AGGTGCTTCGTATGTGTTTGTATTAATTAACCAAGAAAATGAGATTTAAAGCC 1707
OY      482 LeuValAlaArgArgSerTyr---ArgArgArgGlyLysSerGlnGlnGlnGlnGln 500
DB      1708 CTCGTCAAGAAAGGATATTATGGAACCGCAGAGCAAGAAACCAAAATGTCTCCACCTCG 1767
OY      501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
DB      1768 CAAAGAAAGAAAGTAAAGAAAGAAAGAG---GATTAAGCAGAAAGAAACAGAAAG 1821

```

```

OY      521 AspGluLysGluLysGluAlaGluLysGluGluLysProGluValGluAsn 540
DB      1822 AAAGAAAGAAAGAAAGATGAAGAGAGAAAGATGAAAGAAAGAAAGCTCCAAAGAAAT 1881
OY      541 AspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAsnAsnAspGlu 560
DB      1882 ACCAAGGAA-----AAGGACAGATAGATGATACGCA---GAGAAACTGAGGAA 1929
OY      561 LysGluAlaValAlaSerLysArgLysThrAlaAsnSerGlnIysArgArgLysGly 580
DB      1930 AAGAGCAAGACCAACCCCGGGCGGAAAGACTGCAACAGTCAGGCGCGGTAGAGGCG 1989
OY      581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600
DB      1990 CCGATACCAAGGTCCATGACAAAGAGCTGACCGACGTCGACCGCAGCGGCT 2049
OY      601 Ser-----AlaGluLeuAlaSerMetGluLeuAsn 610
DB      2050 ACTGAAGACCCCCACACCTCTGCCACCGCAACAGAACCAATTTCTACAGGCTGTG 2109
OY      611 GluSerSerArgTyrThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluHis 630
DB      2110 GAGACTCTCGATGAGCAGAAAGAAATGGAAGTTGCTAAAGAAAGCTGATGAGAAAT 2169
OY      631 GlyArgAsnTyrSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLys 650
DB      2170 GGTCTGTAATCTGGCGCAATTTCTAAATGTGGGAAACAAAGTCAACCTCAATGTAAA 2229
OY      651 AsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLys 670
DB      2230 AACTCTATTTTAATCTATTAAGGCGACACAAATCTTGACAACTCTTAACGACGCTAAA 2289
OY      671 LeuLysMetGluLysGluLysArgAsnAlaArgArgLysLysLysValProAlaAlaAla 690
DB      2290 CAGAAAACTTCAGAAAACTCGTGAAGAGCAGATGTCTCAATGTGAAGTGCCT 2349
OY      691 SerGluGluAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGlyVal 710
DB      2350 TCCACTGTTTCTGCT-----CAGGAGATGAAGATTTGAACTCTCC----- 2391
OY      711 SerGlyAsnGluGlnGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGlu 730
DB      2392 -----AATGAAGAAATAATCCAGAGACAGGAA----- 2421
OY      731 ValProArgGlyLysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSer 750
DB      2421 ----- 2421
OY      751 IleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProPro 770
DB      2422 -----GTTGAAGCTGTCAAG----- 2436
OY      771 AlaThrLeuGlyAlaAspGlyProProGlyProProThrProProArgThrSer 790
DB      2437 -----CCACGAGAGACAGTCTGAAAAATGTAATCTTCT 2469
OY      791 ArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProProPro 810
DB      2470 CCGAGAAACACAGAACTCGGTGAGCTTGAG-----CCACACACGAGAACT 2517
OY      811 AlaPro---ProSerProSer---AlaProProProValValProLysGluGluLysGlu 828
DB      2518 GCAACCAAGTACATCTCCCTCTTGAAGAGTTCCAAAGTACAAACAGCTGAAGATGAAGT 2577
OY      829 GluGluThrAlaAlaAlaProProVal----- 837
DB      2578 GTGGAGACCCAGGTGAATGACAGCATCACTGCTGAGACACGACAGACATGATGAT 2637
OY      838 -----GluGluGlyGluGluGlnLysProProAlaAlaGluGluLeu 851
DB      2638 CAGCAGAGACACAGTGTGAAGAGGTTCTGTTTGTGATCCCCACCGCTACCAAA--- 2694
OY      852 AlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGlu 871

```

[illegible]

Db	3574	CCCCGAGCCAAACTCAACCTGAGGGTCTGTTGGTC--AGGGCCCAACATGAAAGT	3633
Oy	1194	ValleuATrgtYThAlaLeuglYserValProglYglYserlIethrLYseglYlIePro	1213
Db	3631	GTAGTCAGAGGTACCGCA---GAGCCATACAGAGAGATATATCTCGGGAGACTCCA	3688
Oy	1214	SerThArgrValProSerApsSerAlaIethrTYrArGglYserlIethrHlsglYthr	1233
Db	3688	ACCAACAAATTTTCAGTGGAGAGCATTCCTCAATCCCAAGGGGCTCATCACTCAGGGACAC	3744
Oy	1234	ProAla-----AspValleuTYrLYseglYThrlIethrYg	1244
Db	3748	CCGGGCTCTGCCCCAGACTGGCATACCAACAGAGCTTTGGTGAAGGGGTCATTTCGAGA	3807
Oy	1246	IlelIeglYlAspSerProSerArgrLeuAspArgrYArGglYAspSerleuProlys	1265
Db	3808	ATGCCCATTTGAAGACAGCATGCTCT-----GAGAAAGGCAGAGAGGAAGCTGCATCCAA	3866
Oy	1266	GlYhIseValIleTYrGglYLYseglYhIseValIeSerTYrGglYglYmerSer	1285
Db	3862	GGCCATGTTATTATTAAGAGCAAAAGTGACATATCTTGTCATATGATATATTAAGAAAT	3922
Oy	1286	ValThrGlnCYserLYseglYAspGrYArgrSerSerSerGlyProProhIseglYthAla	1305
Db	3922	GCC-----CGAAGAGGGCTAGAGTCCAGACAGCTCATGAATAATCAGT	3966
Oy	1306	AlAProLYsArgrThrTYrAspMecIeGlYArgrValGlYArGglAlIeSer-----	1322
Db	3967	TTA--AGAGAAAGTATGAATTCAGTGGAGAGAAATATTAAGCAAGGATGTCATTAAGG	4022
Oy	1324	-----SerAlaSerlIeGlYglYLeuMetGlYArGglAlIeProProglYArG	1335
Db	4024	GAGTCTCCTGTATACGACACGTTAGAGGGCTGATATGCGAGACTTATCC-----AGG	4077
Oy	1340	HiserProhIse---HisLeuLYseglYglYhIseIleAlarGglYserlIethrGlnY	1358
Db	4078	GGGAGTCTCTCATCTTGACCTCAAGAAAGAGCTGATATGTGGCTCATATGACAGGGG	4137
Oy	1359	IleProAgsSerTYrValGlYAlaGlYAspTYrLeuArGglValLYseglYLeu	1378
Db	4138	ACACCAAGAGCAACACTGAAGGCTTGAAATGAGCTT---AAATATCCCAACAAATT	4194
Oy	1379	LYsArgrGlYglYThrProProProProProSerArgrAerLeuThrGlYAlATyLYs	1398
Db	4195	AAAAGGAAAT-----CTCTCCATACAGAGATTTGAAGTGGCAATT---	4236
Oy	1399	ThrGlAlaLeuGlYProLeuLYseglYProAlaHlsGlYglYLeuValAlaThrVal	1418
Db	4237	-----ACCAAGAGAAACCA---TATGATGGC---ATCACCACCATC	4272
Oy	1419	LYseglYAlaGlYArgrSerlIeHlsGlYlIeProArgrGlYglYLeu-----	1433
Db	4273	AAAAGAAATGGGGCTTCATTCAATGAGAAATTCCAAGGCAAGATATTTAATCTCAGAAAGT	4332
Oy	1434	ArgrHlsThrProglYleuProLeuAlAProArgrProLeuLYseglYglYserlIethrGln	1453
Db	4333	CGAAAACTCCAGAAAGTGGTCCAGAGCAACAGCGCCGATATTAAGGGTTCATTTCCAG	4392
Oy	1454	GlYThrProLeuLYsTYrAspThrGlYAlaSerThrThrGlsYserLYseglYhIseArVal	1473
Db	4393	GGCAGACCAATTAAGTTTGAACAACAC---TCAGGTCAATCTGCCATCAACACAAATGTC	4449
Oy	1474	ArgrSerleuIleGlYserProglYArgrThrPheProProValHlsProLeuAspValMet	1493
Db	4450	AAATCTTTAATCAAGGGGCTTCAGCAAACTATCCCGTGAATGCTCCGCTGGAATTTGTG	4509
Oy	1494	AlaAsp---AlaArGglAlaLeuGlYArGglAlaCYsTYrGglY-----GlYser	1507
Db	4510	CCAGAGAAACATTAAGTGTGAACCGGGAATAATAGCATGTGAAGCAAGCGGAGAAC	4569
Oy	1508	LeuLYsSerArgrProglYThrAlaSerSerSerGlyglYserlIeAlaArGglYAlaPro	1527
Db	4570	GTGGATTCGCGGCACAGTCAAGTGTGAAGCTGTGGCCCTCGGTTCTTAAGTCCACA---	4626

OY	1528	ValIleValProGluIleuGlyIleuProArgGlnSerProIleuThrTyrGluAspHisGly	1547
Db	4627	-----CTGCAATGAAAGCTCCCAAAAGCAACAGTGAAGCTCGGTGATTATGATGACACAGT	4680
OY	1548	Ala-----ProPheAlaGlyIleuProArgGlySerProValThrMet	1562
Db	4681	GCAAGGAGAGACCCCTGTGATGTTATTAATAACCAATGTCACAAAGGCTCACCATATGTAAC	4740
OY	1563	ArgGluProThrProArgIleuGlnIleuIleuSerIleuSerValSerGlnAsp	1582
Db	4741	AGAACTTGATGATGTTACAAATTCCT-----CCTAACAGTACCAATCATGAA	4788
OY	1583	ArgIleuIleuThrSerThrProArgGlu-----IleAlaIleuSerProHisSer	1598
Db	4789	AGGAATATGACATCGACCCCTTACCCAGGAGGAAGTATCCAGGAAATCTCCAGTCTC	4848
OY	1599	ThrValProGluHisIleProHisProIleSerProTyrGluHisIleuLeuArgGlyVal	1618
Db	4849	GGGTGACCCCTGCTGTAGCCAC-----AGTCCGTTGATCCCATACAGAGCCAGC	4902
OY	1619	SerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIlePro	1638
Db	4903	ACTGAGGCGAGGTTATTTAGAGCCACTGCGCACGCAATGATGATCA---GCCATGCT	4959
OY	1639	ArgGlyIleProLeuAsp---AlaAlaAlaAlaTyrTyrLeuProArgHisIleuAlaPro	1657
Db	4960	TTTCACAGAGGCTTTGGATCTGCAAGCGCTGCTTACCTGTTCAAGACAGCTTCCACA	5019
OY	1658	AsnProThrTyrProHisIleuTyrProProTyrIleuIleArgGlyTyrProAspThrAla	1677
Db	5020	ACTCCAGGTTACCCAGTCAGTATACGCTTAC-----	5052
OY	1678	AlaIleuGluAsn---ArgGlnThrIleIleAsnAspTyrIleThrSerGlnIleMetHis	1696
Db	5053	GCAATGAGAAACACACAGACAGACAACTTAAATGATTCATTACTCCACACAGATGCA	5112
OY	1697	HisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyIleuSerProArg	1716
Db	5113	GTGAACCTTG-----CGTCCAGATGTGGCCAGAGCACTCTCCACAGA	5154
OY	1717	GluSerSerLeuAlaIleuAsnTyrAlaIleGlyProArgGlyIleIleAspLeuSerGln	1736
Db	5155	GAGCAGCCACTGGCTGCCATCCACAGCA---ACGAAGAGATCATTTGACTGACCAAT	5211
OY	1737	ValProHisIleuProValIleuValProProThrProGlyIleProAlaThrAlaMetAsp	1756
Db	5212	ATGCTTCCA---ACAATTTTATGTCCTCATCCAGGGGGAACAAGACATCTCCCATGAGC	5268
OY	1757	ArgLeuAlaTyrLeuProThrAlaProGlnProPheSerArg---HisSerSerSer	1775
Db	5269	AGAACTACTTATATTTCTGCTACACAGATTACTTTCCTCCAGGCGCTGACAACTCTGCT	5328
OY	1776	ProIleuSerProGlyIleProThrHisIleuThrIleuProThrThrSerSerSerGlu	1795
Db	5329	TCCATGCTTCCAGGACACCCACACACTT-----GCAGCTGCTGCAAGTCTGAG	5379
OY	1796	ArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGlyIleuSerIleu	1815
Db	5380	AGGGAACGGGAACGGGAGCGGAGAGAGCGGAGCGGGAACGG-----ATTGCT	5430
OY	1816	ThrSerThrThrThrValGlnHisAlaProIleThrPargProGlyIleThrGlnGlnSerSer	1835
Db	5431	GCAGTTCTCTCCGACCTC-----TACTGGCGGACAGGCTTGAAACAG-----	5472
OY	1836	GlySerSerGlySerSerGlyGlyIleGlyGlySerSerSerArgProAlaSerHisSer	1855
Db	5473	-----CCTGACCGACCTGGCAGTATGCA	5496
OY	1856	HisAlaHisGlnHisSerProIleSerProArgThrGlnAspAla---LeuGlnGlnArg	1874
Db	5497	TATGTGCTGCTCCCTTCCCT---TCACATTAAGAACTCAGAGAGACATGTGTCACACAGAGA	5553

Qy	1875	ProSerValLeuHhIaenThiGlyMetLeuValLeIleThiAlaValGluProSerLeu	1894
Db	5554	CCAGGTGTTTTCCAAAGAACCAATGGAACCAAGTAATACACTTTGGATCCAACTGCT	5613
Qy	1895	ProThiValLeuArgSerThiSerThiSerSerProVal	1910
Db	5614	CAGCTACGAACTATGCCCACTGCTGCTGGGGGCGCTTCAAATACCAAGGGCTGCCAGCC	5673
Qy	1911	AlaThiRheProProValAlaThiHicGyProLeuValGlyThiLeuArgValValTyPro	1930
Db	5674	TCCGGTTACAAACTGCTGGCGAATGCC---CTGGCTGCTTTGGATGCTGCGAGCTTCT	5730
Qy	1931	ThiLeuMetGluProValLeuLeuProLys	1946
Db	5731	GCACCCAGATGATGTGTCTCCAAAACAAAGAAGTAAGCATGAAGCTGCAGGTTGAA	5790
Qy	1947	ArgProGluArgProValAlaAsp	1954
Db	5791	GAAATTTTGAAAGCAGGTGTCAGACAGATTGTAACAGCAGCATGAGACGAAACC	5850
Qy	1955	-----ThiGlyHhIaPheLeuAlaLys	1962
Db	5851	CTGGAAGCTGGAAGAAAGATCTGTTCAGGTTTATACACTTCTTCAAGCTTCCAAATGGC	5910
Qy	1963	ProProAlaArgSerGlyLeuGluProAlaSer	1979
Db	5911	AAGCCC-----CAGCTCATCTTTCAGATGTTATTTCTGAAGCTGGGAA	5955
Qy	1980	GluProAlaArgProLeuValProProValSerGlyHhIaThiLeaIleArgThiProAla	1999
Db	5956	GATAAAGGGCT---CTCCAAATCCAGATATGAGAAAGCTTAAGACCAAGG	6009
Qy	2000	LysAsn-----LeuAlaProHhIaHhIaSerProAspProProAlaProProAlaSer	2017
Db	6010	AAGATACCATTAATCTGAGAGTAACTTACAGCTATCAACCGGCAATATGGCTCG	6069
Qy	2018	AlaSerAspProHhIaArgGlyLysThiGlnSerLysProHieSerIleGlnGluLeuGlu	2037
Db	6070	GACAAAGATGCGAAGGAACTGGCTCTCAAGTTCAACTCTTCT	6114
Qy	2038	LeuArgSerLeuGlyTyHhIaGlySerSerTySerProGlnGlyValGluProValSer	2057
Db	6115	---AGTAGCTTATCTTCTCACAGATGAACAACTAGCAATGCTATTAAGAGTAGTAAGT	6171
Qy	2058	ProValSerSerProSerLeuThiHhIaAspLysGlyLeuProLysHhIaLeuGluLeu	2077
Db	6172	CTCGGCACTCACTGCGCCAGCCAGGAAATGCGACACTTATCAAGCCAGAGTTGTT	6231
Qy	2078	AspLysSerHhIaLeuGluGlyGluLeuArgProLysGlnProGlyProValLysGly	2097
Db	6232	AAGGCAATCAAGCGGAATGATCTTCCAGACAAATGTAAGGACCA-----	6279
Qy	2098	GlyGluAlaAlaHhIaLeuProHhIaLeuArgProLeuProGluSerGlnProSerSerSer	2117
Db	6280	-----TTACATCACTATGACCA-----CAGCAGGAATCACTACT	6315
Qy	2118	ProLeuLeuGlnThiAlaProGly-----ValLysGlyHhIa	2129
Db	6316	CCCCCAACCAAGCTGCCCTTCTTTCACAGGACAGAGGAATGGGCAAGTCCAGAGAC	6375
Qy	2130	GlnArgValValThiLeuAlaGlnHhIaIleSerGluValIleThiGlnAspTyThiArg	2149
Db	6376	CATCGGCTGATCACTTCTGATGATCAATCTGTCAAAATTAACAAAGATTTTGCTAGA	6435
Qy	2150	HhIaHhIa-----ProGlnGlnLeuSerAlaProLeuProAlaProLeuTy	2164
Db	6436	AATCAAGTTTCTCGCAGACTCCCGACAG-----CTCTTACTTTCATCTCCAG	6486
Qy	2165	SerPheProGlyAla-----SerCyProAlaLeuAspLysLeuArgArgProProSerAsp	2182
Db	6487	AACCTACCTTCTGCTTGGATTTACACTGTG-----AGACTAAACATCAAC	6537
Qy	2183	LeuTyTyLeuProProProAsp-----HhIaGlyAlaPro--AlaArgLys	2196

```

Db      6538 CGTTACGACCAGAAATCCAGGCTCAGTCTGTCCATCATCAAGACAGGCTTCAAGGCTC 6597
Qy      2197 SerProHis-----SerGluGlyGlySerAspSerProGluProAsnLys 2211
Db      6598 TCTTCAGAAATCTTGGAGCAAAATCCAGGGAAGTGGCTCGAATAATCCCAAGAGAG 6657
Qy      2212 ThSerValLeuGlyGlyGlyGluAspGly11LeuLuproValSerProProGluGlyMet 2231
Db      6658 AGTACAGTC-----TCTCCGAGCCCTACAGAGCCCATCTCCACCCAG----- 6702
Qy      2232 ThrGluProGlyHisSerAspSerAlaValTyrProLeuLeuTyrAspAspGlyGluGln 2251
Db      6703 GTTCCGCTTGCTGCTGAGAAACAGACAGAGCTTGTCTGTCTGCTACAGAGGGGCGCAGAG 6762
Qy      2252 ThrGluProSerAspMetGlySerTyrSerProGlyLysAsnThrSerGlnProProAlaPhe 2271
Db      6763 CTCGCAAGACAGAGAAATGATGCCCCGCTCACAGAGGATATAGCTTACTTGCCTTCAATC 6822
Qy      2272 PheSerTyrLeuThrGluSerAsnSerAlaMetValLysSerTyrLysGlnGluLysAsn 2291
Db      6823 TTACCCAGAGCT---GAAATACATCACCCATGCTTAATTAATAAGAGCAGAGATTTT 6879
Qy      2292 LysLeuLeuLeuThrHisAsnAspArgAsnGluProGluTyrAsnLysSerGlnProGlyThr 2311
Db      6880 CGTAAGTTGAATCTCTGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 6939
Qy      2312 GluIlePheAsnMetProAlaIleThrGlyThrGlyLysMetThrTyrArgSerGlnAla 2331
Db      6940 GAGATCTTATATCTGCAAGACGATTAATCTAGCAGCTCAGCTTACCTTGAAGGCCATCT 6999
Qy      2332 ValGlnGluHisAlaSerThrAsnMetGlyLeuGluLysAlaIleLeuArgLysAlaLeuMet 2351
Db      7000 TTGCTGATCTCTGCACT---AATCTTGGCTGGAAGACATTAACAGAGAGCTTCTCATG 7056
Qy      2352 GlyLysTyrAspGlnTyrGluGln-----SerProProLeuSerAla 2365
Db      7057 GGAAGCTTATGATGCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 7116
Qy      2366 AsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAla 2385
Db      7117 -----GTGCTGCTGACTGCTCAACACCTCACT-----GTGACCACT 7152
Qy      2386 AlaAspGlyArgSerAspHisIleThrLeuThrSerProGlyGlyGly-----LysAla 2403
Db      7153 GGTGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7212
Qy      2404 LysValSerGlyArgProSerSerArgLysAlaLysSerProAlaPro-----GlyLeu 2421
Db      7213 AAGCTGATCAGCAAGTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7272
Qy      2422 AlaSerGlyAspArgProProSerValSerSerValHisSerGlyGlyAspCysAsnArg 2441
Db      7273 TTAGGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7332
Qy      2442 ArgThrProLeuThrAsnArgValTyrGluAspArgProSerSerAlaGlySerThrPro 2461
Db      7333 CAAACCGCA-----GGGTGGGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7386
Qy      2462 PheProTyrAsnProLeuLysMetArgLeuGlnAlaGlyValMetAlaSerProProPro 2481
Db      7387 TTTCCTTATTAACCTCTGATGATGCGGATG-----CTCAGCAGTACTCCACCAACA 7437
Qy      2482 ProGlyLeuProAlaGlySerGly-----ProLeuAlaGlyProHis-----HisAla 2497
Db      7438 CCAATTGTCATGTCCTCTCTGCGGTGAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGATC 7497
Qy      2498 TTPAspGluGluProLysProLeuLeuCysSerGlnTyrGluThrLeuSerAspSerGlu 2517
Db      7498 TGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 7557

```

RESULT 15
AF087856

pol1a_signal

```

LOCUS      AF087856      8018 bp      mRNA      linear      PRI 23-JUN-2003
DEFINITION Homo sapiens nuclear receptor co-repressor mRNA, complete cds.
ACCESSION  AF087856
VERSION    AF087856.1  GI:33150575
KEYWORDS
SOURCE     ORGANISM
            Homo sapiens (human)
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 8018)
            Tu, Q., Yu, L., Dai, F., Yue, P., Huang, H. and Zhao, S. Y.
            Cloning and characterization of a new human cDNA homologous to
            murine nuclear receptor co-repressor mRNA
            Unpublished
            2 (bases 1 to 8018)
            Zhang, M., Yu, L., Zhou, Y., Hu, P. R., Xin, Y. R. and Zhao, S. Y.
            Direct Submission
            Submitted (27-AUG-1998) Lab of Human Gene Research, Institute of
            Genetics, Fudan University, No. 220 Handan Rd., Shanghai 200433,
            P. R. China
FEATURES
source
CDS
FEATURES
location/Qualifiers
1..8018
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
85..7116
/note="nucor"
/codon_start=1
/product="nuclear receptor co-repressor"
/protein_id="AAP97166.1"
/db_xref="GI:33150576"
/translacion="MSSVYPNOCAGAFSTEQSRSPPHSVKTEPSTTHOODPAFGGHI
EAPSPISGCGDDONASPSKISKEILDSMDRVREIAKVEQILKIKKQOOLE
BAKPPPEPEKVPSPVEQKHSTIVQIYDNRKDAEAKITEGAGPKYELPLYNP
SDTKVHTENKTNQVNRKKILFKRNHAKRQEQKICQITQULWEAKKVDRIEN
NPRKAKSEKTRREYKQPEIRKQEOQERFQVQGRGLSATIARSHEISEIID
GLSEOSNEKOMQLSVIPWMDACORRYKFINMGGLMDPMKVVADROFMMWTDH
EKEIFDKFIQHPKNGFIASLYLERISVPCVLYLTIKENYKALVRNYSRGR
NOQIARPOSEKTEKEEDKAEKTEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEE
ETBERQATPRGKTKANSQCRKGRITRTSTNTAAAAASAAAAATEPPPLPPPEP
ISTEPVETSKWTEEMEVAKGLVEGRNMAALAKVGTSEAKQKFNFYKRNL
DNLQHQKQKTSKPREBVDQSESVASVNAQEDBDLEASNEENBEDSEAEANS
DIESAPSPVEAIVKPSDESPENATGRGTEPVELEPTEPTAPSTSLAVSTKA
EDESVETQVNDISISETAEDMDVDQOEHSEBSGVCDPPATADSDVVRPENNA
SKYEGNTKRDRLDRASEKVEPRDELDVAQOINARBPSONDSATCSADEVDG
EPRQCMFPRDSCRSLNPRGSLIVSSPLKRNPLDLPOLHRAAVIPMYVSCPCNIP
IGTVSGVALYQKHITAMHSALILEQROBOIDLECRSTSPCGTSKPNRBEVL
QPAHQVITNLPBGVRLPTTPRPPLPISKRTIVASEKSPFINGSTISQGTGY
LTSNHOASTQETPRKPSVGSISLGLPROEASAKATLPYIKQEFSPRSQNSPBGIL
VRAOHGCVRGATGALIOEGSITRGTPSKISVSIPLRGSITQGPALPQTGPTEA
LVKGSISMPDIENDSPEKREBAASKGHVYIEGSGHIIISYDNINKNAREGTRPTRAH
EISLKSYSVEBNITKQMSMRESPIYAPLEGICALLPRGSHSLDKERTVYSGSIM
QCTPRATVESFEDGLTYPKQIKRESPIRAFEALITKGAYDQITTIKEMGRHIEIP
RODILIOERKPKTEVVQSTRPIEGSIQOTPIKFNNSGOASIKINVSLITGSPSL
SRGMPLELIVPENIKYVERKGYEDVAKGTVSRHSTVSVSGSPVSLTLEAPKQL
SPGIVYDTSABRTIPVGYONTMRSQPMARSTSVSNKSGNHRKSTLPTPORESIPA
KSPTEGVVDVSHSPDPHHRGSTAGEVYRSLPTLDDMAPRHALDPAAYLPRO
LETPPEYQYQIYMENTROTILANDYTTSQOQVNLRLDPAVAGLSPREQDLPYPA
TRGIDLTNMPPTILVHPFGTSTPMDITTYIPGQITPFPVIVASNSPOHPTLU
AAASAEERERERERERERERERERERERERERERERERERERERERERERERERERERER
ETMLQORPSVFOGTNGTSVITPLDPAQRIIMPGLPAGSPSISQGLPASFRTADALAA
ALVDAASAPOMDVSKTESKHEAALIEULRERSAVSEQOOLEKSTLEKEVERVQC
LYTSSAPSPSGKQPPHSSVYVSEKQKGPSPKRSYREELRTTRKTTITANFIDVIT
ROIASDKDAREGSSQSDSSLSHSRYETPSDAIVIVSPASPAPQELQYQYQEV
VKANQENPPTROYEGPLHRYRPOQSPSPQOOLPVSQOQNGOVPRTHRLTLTLDH
ICQITQDPAFRQVSSQTPQOPEPTSTFQNSPSALIVPTKRTSNVSPESQOQVHH
ORPSRVSPEBNLVDSKPSRPGSPERSHVSSEPPYIPBPQPVVHAKQDLSLLSQ
ORGPALQORNDNAPSGISYLPSPFKLENTSPMYSKQELFRKLNSGGGSDSMAA
AOGGTIFNLPAVTTSVSGSHSPADPSNGLBIDIRKALMGFDDKVEDHGVVM
SOPMGVPRNTATSVTSGSTREREDPSPHSGVQKDKLISKNSRKSQSPIDGQGLV
GTRPSVSVSHSBDYHROTQPMANEDPSSGTSQTFPYNPLTMMSLSTPPTIAC
APSAVQAAPHOQNRIMEREPAVLISAQYETLSDSD"
7984..7989

```

ORIGIN

Alignment Scores:

Pred. No.:	1,396-73	Length:	8018
Score:	4043.00	Matches:	1061
Percent Similarity:	54.21%	Conservative:	330
Best Local Similarity:	41.35%	Mismatches:	731
Query Match:	30.59%	Indels:	444
DB:	9	Gaps:	104

US-09-522-753-5 (1-2517) x AF087856 (1-8018)

```

QY 137 LysAspArgSerLeuThrGlyLeuGlu---ProValSerProProSerProProHis 155
DB 190 CAGGATCCAGATTCGAGGCAAAACATGAAGCTCCATCTCTCAATTCGGGGCAACCA 249
QY 156 ThrAspProGluLeuGluLeuValProProArgLeuSerLeuGluLeuLeuLeu 175
DB 250 TGTGAGATGATCAAAATGCTTCACCTTCACAACTCTCAAGAGAGATTAAATACAGAGT 309
QY 176 MetAspArgValAspArgGluLeuThrMetValGluGluGlnLeuSerLeuLeuLeu 195
DB 310 ATGCGATCGTGATGATGAGAAATTCGAAAAGTAGAACAGCATCTTAAACTGAAAAAG 369
QY 196 LysGlnGlnGlnLeuGluGluGluValAlaAlaLysProProGluProGluLysProValSer 215
DB 370 AAACAACAACAGCTTGAAGAAGAGCAGCTAAACCTCTGAGCCTGAGAGAGCCCGTCC 429
QY 216 ProProProGluLeuSerLeuSerLeuValGlnLeuLeuLeuLeuLeuLeuLeu 235
DB 430 CCTCCCTCTGTGAGCAAAACCCGAGATTTGTCCAAATTTATGATGAGAAATCGG 489
QY 236 LysLysAlaGluAlaAlaHisArgLeuLeuGluGluGluProGluValGluLeuPro 255
DB 490 AAAAAAGCAGAAAGACTCTATAAATTTTGAAGCTTGGCCCAAAAGTGAACCTGCCA 549
QY 256 LeuTyraGlnGlnProSerAspThrArgGlnTyraGlnAlaGlnLeuSerLeuGlnAla 275
DB 550 CTGTATTAACAGCATCAGATACCAAGGTGACCATGAGAACATCAAGCAAAACCAAGGTG 609
QY 276 MetArgLysLeuLeuLeuLeuTyraPheLysArgArgAsnHisAlaArgLysGlnTyrLys 295
DB 610 ATGAGGAAAAAACTCTATTTATTTTAAAGAAAGAAATCATCTCAAGAAAAACAAAGGAA 669
QY 296 GlnLysPheCysGlnArgTyraAspGlnLeuMetGluAlaLeuGluLysLysValGluArg 315
DB 670 CAATAAATCTGCCAGCGTTATGATCACTCATGAGGATGGAGAAAAAGTGGACAGA 729
QY 316 IleGluAsnAsnProArgArgArgAlaLysGlnSerLysValArgGluTyraGluLys 335
DB 730 ATAGAATAATATCTCTCGAGAGAAAGCTAAAGAAAGCAAAACAAGGAAATCTATGAAAAG 789
QY 336 GlnPheProGluLeuLeuArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGly 355
DB 790 CAGTTTCCAGAAATTCGAAAAACAAGAACACCAAGAAAGATTTTCAG---CGAGTTGGG 846
QY 356 GlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluLeu 375
DB 847 CAGAGGGAGAGCTGCTTTTCAAGCCACCATTTGCTAGAGTGAAGCATGAGATTTCTGAAAT 906
QY 376 IleAspGlyLeuSerGluGlnGluLeuValLeuGlnLysGlnMetArgGlnLeuAlaLeu 395
DB 907 ATTGATGGGCTCTCTGAGCAGGAAGTAATGAGAAACAATTCGGCGAGCTCTGTGATTT 966
QY 396 ProProMetLeuTyraAspAlaAspGlnGlnArgGlnLeuLeuPheLeuLeuMetAsnGlyLeu 415
DB 967 CCACTATATGATGTTGATGCAAGAACAAAGCAGATCAAGTTCAATTAACATGAATGGGCTT 1026
QY 416 MetAlaAspProMetLysValTyraLysAspArgGluValMetAsnMetTyrSerGluGln 435
DB 1027 ATGAGAGACCTTATGAAAGTGTATTAAGATAGGCACTTTATGAAATGTTTGGACTGACCAT 1086
QY 436 GlnLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuLeuAla 455

```

```

DB 1087 GAAAAAGAGATCTTTAAGACAAAGTTTATCCAGCATCCAAAAAAGCTTGGACTAATTGCA 1146
QY 456 SerPheLeuGluArgLysThrValAlaGluCysValLeuTyraTyraLeuThrLysLys 475
DB 1147 TCATCTTGGAGAGAGAGAGTCTTCTGATGTGTTTGTATTTATTCATTTAAACCAAGAAA 1206
QY 476 AsnGluAsnTyraLysSerLeuValArgArgSerTyr---ArgArgArgGlyLysSerGln 494
DB 1207 AATGAGAAATTTAAGAGCTCTGTCAGAAAGCAATTAAGGAAACGACGAGCGAGAAACAG 1266
QY 495 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 514
DB 1267 CAATATTGCTCGACCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
QY 515 SerSerGlnGlnGluLysAspGlyLysGluLysGluLysGluLysGluLysGluLysGlu 534
DB 1321 GCAGAAAAAACAAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
QY 535 LysProGluValGluAsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSer 554
DB 1381 GAAGACTCCAAAGAAAAATACCAAGGAA-----AAGAGACAAGTATGATGTCACAGCA 1431
QY 555 GlyLysAspAsnAspGlyLysGluValAlaValAlaSerLysGlyArgLysThrAlaAsnSer 574
DB 1432 ---GAAGAAACTGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1488
QY 575 GlnGlyArgArgLysGlyArgLysLeuThrArgSerMetAlaAsnGluValAsnSerGluGln 594
DB 1489 CAGGCGCCCGCTTAAGGGCCGGATCACAGGTCCATGACAAAGAGAGCTCAGCTGCACGT 1548
QY 595 AlaIleThrProGlnGlnSer-----AlaGluLeu 604
DB 1549 GCTGCGAGCCGACGCGCTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1608
QY 605 AlaSerMetGluLeuAsnGlnSerAspArgTyrThrGluGlnGluMetGluThrAlaLys 624
DB 1609 ATTTCTACAGACCTGTGAGAGACTCTGATGAGCAAGAAAGAAATGGAAGTGTCTAAA 1668
QY 625 LysGlyLeuLeuGluHisGlyArgAsnTyrPheSerAlaAlaAlaArgMetValGlySerLys 644
DB 1669 AAGAGCTTGTAGAACATGTGTCTTAATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1728
QY 645 ThrValSerGlnCysLysAsnPheTyraPheAsnTyraLysLysArgGluAsnLeuAspGlu 664
DB 1729 AGTGAAGCTCAATGTAAAACTTTATTTAATTAATAAGAGAGAGAGAGAGAGAGAGAG 1788
QY 665 IleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLys 684
DB 1789 CTCTTACAGCAGCATTAACAGAAACTTCACAGAAACCTCGTGAAGAGAGAGAGAGAGAG 1848
QY 685 LysAlaProAlaAlaAspSerGluGluAlaAlaPheProProValAlaGluAspGluGln 704
DB 1849 CATGTGAAGAGTGTCTTCCACTGTTCTGCT-----CAGAGAGATGAAGAT 1896
QY 705 MetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGluAlaLeu 724
DB 1897 ATTGAAGCTCC-----AATGAAGAAAGAAATCCAGAGAGAGAGAGAGAGAGAG 1941
QY 725 HisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsn 744
DB 1942 -----GCAGAAAT 1950
QY 745 SerSerAspThrGluSerLysProSerProHis---ThrGluAlaAlaLysAspThrGly 763
DB 1951 AGTTCTGATACAGAAAGTGTCTCTTCTTCTTCAACAGTTGAAGCTTCAAG----- 2001
QY 764 GlnAsnGlyProLysProProAlaThrLeuGluAlaAspGlyProProProGlyProPro 783
DB 2002 -----CCAGAGAGAGAG 2013
QY 784 ThrProProArgArgThrSerArgAlaProIleGluProThrProAlaSerGluAlaThr 803

```


Db	2014	AGTCTGAAAATGCTACTTCTCGAGAAAACACAGAACTCGCGGTGAGCTTGAG-----	2067
Qy	804	GLyAlaPProthrProProProAlaPro--ProSerProSer--AlaProProProVal	821
Db	2068	-----CCACCAACGAGAAACCTGCACCCAGTACATCTCCCTCCTTAGCAGTTCCAAGTACA	2121
Qy	822	ValProlyBSGIuGIuLyBSGIuGIuThraAlaAlaAlaProProVal-----	837
Db	2122	AAACCACTGAAGATGAAGAAAGTGTGGAGACCCAGGTGAATGACAGCATAGTGTGAGACA	2181
Qy	838	-----GluGIuGIuGIuGIuLyBS	844
Db	2182	GCAGAGCAGATGATGATGATCAGACAGACAGCAGTGCCTGAAGAGGGTCTGTTGTGAT	2241
Qy	845	ProProAlaAlaGIuGIuLeuAlaValaAlaPThrGIuLyBSAlaGIuGIuProValLySer	864
Db	2242	CCCCCAACCCCGTACCAA-----GCTGACTGTGAGACGTGTAAGTAGAGGTGCACAGA	2295
Qy	865	GIuCySerThrGIuGIuAlaGIuGIuGIuGIuProAlaLyBSGIuLyBSAlaGIuAlaGIu	884
Db	2296	AACCATGCATCTAAAGTTGAAGGTGATTAATACCAAGAAAGAGACTTGGATGAGACC---	2352
Qy	885	AlaThrAlaGIuGIuAlaLeuLeuAlaGIuLyBSGIuLyBSGIuGIuLySerGIuYrAlaIaThr	904
Db	2353	-----AGTGAAGAGTGGAACTCAGAGATGAAGATTGGTG	2388
Qy	905	ThraIaLySerSergIyAla-----ProGln---AspSerAspSerSerAla	919
Db	2389	GTAGTGTACGAAATTAATGCCCAAGGCCGACGCCGACTGCAGACAAATGATTCCAGTGC	2448
Qy	920	ThrCySerAlaAspGIuValaAspGIuAlaGIuGIuGIuYrAlaAspLyBSAlaIyLeuLeuSer	939
Db	2449	ACGTGCAGCGCTGATGAG-----GATGTGATGAGAGAGCCAGAGAGCAGACGATGTT	2502
Qy	940	Pro-----ArgProSerLeuLeuThrProThrGIuYrAspProYrAlaAlaAlaSer	956
Db	2503	CCTATGCAGCTCAAGCCTTCACTGTAAACCCCACTGGATCTATATCTGTC--TCATCT	2559
Qy	957	ProGlnLyS-----ProLeuAspLeuLyBSGlnLeuLySGlnYrAlaAlaAlaIlePro	974
Db	2560	CCGTTAAAAACCAATCCACTGATCTGCGACAGCTTCAGCATCGAGCTGCTGTTATCCA	2619
Qy	975	ProIle-----	976
Db	2620	CCATGATCTCTGCACCCCATGTAAACATACCAATTGGAAACCCAGTGAGCGCTATGCT	2679
Qy	977	-----GlnValThrIyValAlaIleGIuProProAlaYrGIuAlaAlaIaProThr	992
Db	2680	CTTACACAGCAGACATTAAAGACATGATAGTCAGCATCTCTGAGAC-----	2727
Qy	993	LySProAlaProProAlaProProProGlnAlaSerGlnProGlnLySerAspAlaPro	1012
Db	2728	-----GAGCAGCGCGCAGAGACAAAGAACAGATGATTTGAA	2763
Qy	1013	GlnGlnProGIuLySerSerProArgIyLyBSerAlaYrSerProAlaProProAlaAspLyS	1032
Db	2764	TGTGAAGATTCTCAAGATCGATGTGGACATGCCAAGAGTCCA-----AACAGA	2811
Qy	1033	GIuAlaIaPheAlaAlaGIuAlaGlnLyLeuProGIuYrAspProProCySerThrThSerGIy	1052
Db	2812	GAG-----	2817
Qy	1053	LeuProPheProValProProArgGIuValIleLyBSAlaSerProHAlaAlaProAspPro	1072
Db	2818	-----GAACTCTTGACCTGCTCCACAT-----	2841
Qy	1073	SerAlaPheSerIyAlaProProGIuYrAlaProLeuProLeuGIuLeuHAlaAspThrAla	1092
Db	2842	CAAGTGATTAATAATCTCCGGAAGGGCTGGCGTTCCG-----ACAAC	2886
Qy	1093	ArgProValLeuProAlaYrProProThrIleSerAlaProProProLeuIleSerSerAla	1112
Db	2887	CGACCAACCAAGCCACCGCCCTCTCATCCCGTATCCCAAAACCAACAGCTTGTCAGA	2946

QY	1113	lyshhAPrProSerValIleuGIuAArgGlnIleGIyAlaIleSerGlnGly	-----	1128
Db	2947	AAA---CCATCTTTTATTA-----ATGGAGAGGCTCCATCTCAGGGAAACACGAGCACT		2997
QY	1139	--MetSerValGIuIleuHleValIProIyZSerGIuHlsAlaIyAlaPro--ValGIy		1146
Db	2998	TATTTAGCTTCTGCATTAATCAGAGCTTCCTAGCATCAGAAACACCCAGGCGTCAGTGGGA		3057
QY	1147	ProValThrMetGIyLeuProIeuProMetAspProIySlyIleuAlaProPhSerGIy		1166
Db	3058	TCTATCTCTCTTGGAGCTGCCACGCGCAACAGAAATCTGCCAAATCAGCTTACTTGGCCCTAC		3117
QY	1167	ValIySgInGIuGIuIleuSerProArgGIyGlnIleGIyProProGIuSerLeuGIyVal		1186
Db	3118	ATCAAGCAGAGAAATTTTCTCCCCGAGACCAAAATCACAACCTGAGGCTCTTGGTGC		3177
QY	1187	ProThrIaGIuGIuAlaSerValIleuArgGIyThrIaIleuGIySerValProGIyGIy		1206
Db	3178	--AGGGCCCAACATGAAGGTGTGATCAGAGTCCGCA---GGAGCCATTCAGAAAGGA		3231
QY	1207	SerIleThrySgIyIleProSerThrArgValProSerAspSerAlaIleThrTyArg		1226
Db	3232	AGTAAATCTCGGGAACTCCCAACAGCAAAATTTCACTGAGACATTCATCCCTTAGG		3291
QY	1227	GlySerIleThrHlsGIyThyProAla-----AspValIleu		1238
Db	3292	GGCTCTATCATCATCAGGGACCCCGGCTCTGGCCAGACTGGCATACCAACAGAGCTTTG		3351
QY	1239	TyrIySgIyThrIleThrArgIleIleGIyGIuAspSerProSerArgIleuAspArgGIy		1258
Db	3352	GTGAAGGGGTCCATTTGCAGAAATCCCATCTGAACAGACAGACTCT-----GAGAAAGGC		3405
QY	1259	ArgGIuAspSerIleuProIySgIyHlsValIleTyGIuGIySlySgIyHlsValIleu		1278
Db	3406	AGAGAGGAAGCTGCATCCAAAGGCCATGTTATTATAGCCAAAAAGTGACATATCTTG		3465
QY	1279	SerTyGIuGIyGIyMetSerValThrGlnCySerIySgIuAspGIyArgSerSerSer		1298
Db	3466	TCATATGATATATTATTAAGAAATGCC-----CGAGAAAGGACTACGAGTCCCA		3510
QY	1299	GIyProProHlsGIuThrIaIleAlaProIyAspArgThrTyArgMetMetGIuGIyArgVal		1318
Db	3511	AGAAACAGCTCTTGAATATCAGTTTA--AAGAAAGCTATGAATCATCGTGAAGAAATATA		3567
QY	1319	GIyArgAlaIleSer-----SerIaSerIleGIyGIyLeuMetGIy		1332
Db	3568	AAGCAAGGAGATCTCAATAGAGGAGATCTCTGTATCACACCGCTTAAGAGGGCTGATATGC		3627
QY	1333	ArgAlaIleProProGIuArgHlsSerProHls--HisIleuIySgIuGlnHlsHisIle		1351
Db	3628	CGACACATTAACCC-----AGGGGAGAGTCTCATTTGACCTCAAGAAAGAGACTGTATGG		3681
QY	1352	ArgGIySerIleThrGlnGIyIleProArgSerTyValGIuIaGIuGIuAspTyLeu		1371
Db	3682	TCTGGCTTCATATAGCAGGGAGACCAAGAGCAACAATAAGCTTTGAAGATGGCTT		3741
QY	1372	ArgArgIuAlaIySlyLeuIySgIyArgGIyGIyThyProProProProProSerArg		1391
Db	3742	---AAATATCCCAAAATTAATAAGGAAGT-----CTCCCATTCAGA		3783
QY	1392	AspIleuThrGIuAlaTyTyIyThyGlnAlaIleuGIyProIeuIySlyIleuProAlaHis		1411
Db	3784	GCATTTGAAGGTGCATT-----ACCAAGGAAACCA--TAT		3819
QY	1412	GluGIyLeuValAlaIleThyValIySgIuAlaGIyArgSerIleHlsGIuIleProArgGIu		1431
Db	3820	GATGGC---ATCACACACATCAAAAGAAATGGGGGCTTCCATTCATAGATTCACAGGCA		3876
QY	1432	GIuIleu-----ArgHlsThyProGIuIleuProIeuAlaProIyProIeu		1446
Db	3877	GATATTTTAACTCAGAAAGTGGGAAATCTCCAGAAATGGTGTCCAGAGACAGCGCCGATA		3936

```

OY 1447 LysGluGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThr 1466
DB 3937 ATTAGAGGTTCCATTCCACAGGCGACACCAATAAAGTTGACAAACAAC---TCAGTCA 3993
OY 1467 GlySerLysLysHisAspValArgSerLeuIleGlySerProGlyLysThrPheProPro 1486
DB 3994 TCTGCATCAACAACATGTCAAAATCTTAATCACCGGCGCTGACAAACTATCCCTGGA 4053
OY 1487 ValHisProLeuAspValMetAlaAsp---AlaArgAlaLeuGlnLysArgLysCysTyrGlu 1505
DB 4054 ATGCTCTCCCTGGAAATGTGGCCAGAAACATPAAAGTGTAGAACGGGGAATAATGTAG 4113
OY 1506 -----GluSerLeuLysSerArgProGlyThrAlaSerSerGlyGly 1520
DB 4114 GATGTAAAGCAGCGACGACCGTGCCTCCCGCACAGTCAAGTGTAGCTCTGGCCCC 4173
OY 1521 SerIleAlaArgGlyAlaProValIleValProGlnLeuGlyLysProArgLysSerPro 1540
DB 4174 TCCGTTCTTAGGTCACA-----CTGCATGAAGCTCCCAACACAACTGAGCCCT 4224
OY 1541 LeuThrTyrGluAspHisGlyAla-----ProHisAlaGlnHisLeuPro 1555
DB 4225 GGGATTATGATGACCCAGTGCACGAGAGACCCCTGTAGATTATGAAAACACCATGTC 4284
OY 1556 ArgGlySerProValThrMetArgGluProThrProArgLeuGlnGlySerLeuSer 1575
DB 4285 AGAGGCTCAACCATGATGAACAGA-----ACTTGATGATGTTCT 4323
OY 1576 SerSerLysAla---SerGlnAspArgLysLeuThrSerThrProArgLys----- 1591
DB 4324 TCTAACAACTCAACCAATCATGAAAGAAATGCACACTGACCCCTCCAGAGGGAAGT 4383
OY 1592 ---IleAlaLysSerProHisSerThrValProGlnHisHisProHisProLysSerPro 1610
DB 4384 ATCCACGCGAAGTCTCAGTGCCTGGGGTGGACCTGTGTGACCCAC-----ATTCGG 4437
OY 1611 TyrGlnHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisValProLeu 1630
DB 4438 TTTGATCCCATCCACAGGACGACACTGCAGCGAGATTATGGAGCCACCTGCCACG 4497
OY 1631 AlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaIleTyr 1650
DB 4498 CACTTGATGCA---GCCATGCTTTTTCACAGGGCTTTGGATCCTCGGTGTACTCG 4554
OY 1651 LeuProArgHisLeuAlaProAspProThrTyrProHisValLeuTyrProTyrLeuIle 1670
DB 4555 TTTCAGAGACGCTTTCACCACTCCAGGTTACCAAGTCAGTACAGCTTAC----- 4608
OY 1671 ArgGlyTyrProAspThrAlaAlaLeuGluAsn---ArgGlnThrIleIleAsnAspTyr 1689
DB 4609 -----GCCATGGAGAACACAAAGACAGACATCTTAATGATTAC 4647
OY 1690 IleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMet 1709
DB 4648 ATTACCTCAACACAGATGCAAGTGAACCTG-----GCTCCAGATGTG 4689
OY 1710 LeuArgGlyLeuSerProArgGlyLysSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArg 1729
DB 4690 GCCAGAGACTCTCCCAAGAGACGACCACTGGGTCCTCCATACCCAGCA---ACGAGA 4746
OY 1730 GlyIleIleAspLeuSerGlnValProHisValLeuProValLeuValProProThrProGly 1749
DB 4747 GGAATATTGACTGACCAATATGCGCTCA---ACAATTTTACTGGCTCATCCAGGGGGA 4803
OY 1750 ThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSer 1769
DB 4804 ACAAGCACTCTCCCAAGACAGAAATCACTTATATCTCGTACACAGATTACTTCCCT 4863
OY 1770 SerArg---HisSerSerProLeuSerProGlyGlyProThrHisLeuThrLysPro 1788
DB 4864 CCCAGCCGAGCAACTCTGCTTCATGTCCTCCAGACACCAACACACCTT----- 4914
OY 1789 ThrThrThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspArg 1808

```

```

DB 4915 GCAGCTGTGCAAGTGTGAGAGGAGGAGAACCGGAGCGGAGAGAGCGGAGCGG 4974
OY 1809 GlnArgGlnLysSerIleLeuThrSerThrThrValGlnHisAlaProIleTyrArg 1828
DB 4975 GAACGG-----ATTGCTGACACTTCTCCAGCTC-----TACCTGGCG 5013
OY 1829 ProGlyThrGlnGlnSerSerGlySerSerGlyGlyGlyGlyLysSer 1848
DB 5014 CCAGGCTCAGAACAG-----CCT 5031
OY 1849 SerArgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGln 1868
DB 5032 GGCACACTGGCAGTCAATGATGTGCTCCCTCCCTCCCT---TCAGTAAAGACTGAG 5088
OY 1869 AspAla-----LeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIle 1887
DB 5089 GAGACCAATGTTCACACAGACCCAGTGTTTCCAGAGAACCAATGAAACCAAGTATATC 5148
OY 1888 ThrAlaValGluProSerLysProThrValLeuArgSerThrSerThrSerSerProVal 1907
DB 5149 ACACCTTTGGATCCAACCTGACGTACGAATCATGCCACTGCTGGGGGCTTCA 5208
OY 1908 -----ArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGly 1923
DB 5209 ATAAAGCAAGGCTGCGACCTCCCGTATCAACACTGCTGGCGAGTCC---CTGGCTGCT 5265
OY 1924 ThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuProLys----- 1940
DB 5266 CTGGTGATGTGTCAGCTTCTGACCCCAAGATGATGTGTCCAAACAAAGAGAGTAAAG 5325
OY 1941 ---GlnAlaProArgValAlaArgProGluArgProArgAlaAsp----- 1954
DB 5326 CATGAAGTGCAGAGTTAGAAAGAAATTGAGAACAGAGTGCAGCAGTGTAGTGCACAG 5385
OY 1955 -----Thr 1955
DB 5386 CAGCAGCTAGACAGAAAACTGTGAGTGAAGAAAGATCTGTCAAGTGTATACACT 5445
OY 1956 GlnHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGlnProAlaSerSer--- 1974
DB 5446 TCTTCAAGCTTTCACAGTGGCAAGCC-----CAAGCTCATTTCTTCAGTA 5490
OY 1975 -----ProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAla 1992
DB 5491 GTTATTTGAGAGCTGGGAAAGATTAAGGCT-----CTTCAAAATCCAGATATGAG 5544
OY 1993 ThrIleAlaArgThrProAlaLysAsn-----LeuAlaProHisHisAlaSerProAsp 2010
DB 5545 GAAGAGCTTAAGGACAGAGGAAAGCTACCATTAAGTCAAGTCACTTCAATAGACGTGATC 5604
OY 2011 ProProAlaProProAlaSerAlaSerAspProHisArgGlyLysThrGlnSerLysPro 2030
DB 5605 ATCACCCTGGAATGCTCTCGACACAGATGCGAGAGAAACGTGCTTAAAGTTCAGAC 5664
OY 2031 PheSerIleGlnGlnLeuGlnLeuArgSerLeuGlyTyrHisGlySerSerTyrSerPro 2050
DB 5665 TCTTCT-----AGTAGCTTATGTTTCAACAGATGAAACACTAGC 5706
OY 2051 GlnGlyValGluProValSerProValSerSerProSerLeuThrHisAspArgGlyLeu 2070
DB 5707 GATGCTATTGAGGTGATTAAGTCTCTGACACTGACCTGCGCCACCCAGAGAAATGTCAG 5766
OY 2071 ProLysHisLeuGlnLeuLeuAspLysSerHisValLeuGlnGlyGlyLeuArgProLysGln 2090
DB 5767 ACCTATACCCAGAGGTTGTTAAGGCMAATCAAGCGGAAAGATGCTTACAGACATAT 5826
OY 2091 ProGlyProValLysLeuGlyGlyGlnAlaAlaHisLeuProHisValLeuArgProLeuPro 2110
DB 5827 GAAGAGCA-----TTACATCACTATGAGCA----- 5853
OY 2111 GluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGly----- 2125

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 11, 2005, 09:34:57 ; Search time 24239 Seconds
(without alignments)
3952.629 Million cell updates/sec

Title: US-09-522-753-5
Perfect score: 13215
Sequence: 1 MSGSTQLVAQWTRATEPRYP.....WDEPRKPLCSQYETLSDSE 2517

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame_plus.p2n.model -DEV=x1p
-Q=/cgn2.1/USPTO.spool_p/US09522753/runat_02112005_173640_20807/app.query.faeta.1.2695
-DB=EST -QFMT=faeta1p -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blonum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09522753 @CGN 1.1.1005 @runat_02112005_173640_20807 -NCPU=6 -ICPU=3
-NO MAP -LANG=QUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST.*
1: gb_est1.*
2: gb_est2.*
3: gb_hic.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gss1.*
9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	8226	62.2	7372	9	AY412686	AY412686 Homo sapi
2	7109.5	53.8	7013	9	AY412688	AY412688 Mus muscu
3	6533	49.4	6836	9	AY412687	AY412687 Pan trogl
4	2566	19.4	2000	3	BC033087	BC033087 Homo sapi
5	2264	17.1	1362	3	BC020427	BC020427 Homo sapi
6	1643.5	12.4	1966	3	BC026028	BC026028 Homo sapi
7	1634	12.4	1850	3	BC035748	BC035748 Homo sapi
8	1626.5	12.3	1835	3	BC029627	BC029627 Mus muscu
9	1624.5	12.3	1819	3	BC062162	BC062162 Mus muscu

10	1624	12.3	1093	5	BM909096	BM909096 AGENCOURT
11	1603	12.1	1996	3	BC026623	BC026623 Mus muscu
12	1559	11.8	965	4	BM555371	BM555371 AGENCOURT
13	1540	11.7	1126	4	BM553310	BM553310 AGENCOURT
14	1480.5	11.2	1057	4	BM477568	BM477568 AGENCOURT
15	1479.5	11.2	1051	4	BM423558	BM423558 AGENCOURT
16	1471.5	11.1	937	5	BQ949384	BQ949384 AGENCOURT
17	1448	11.0	846	5	B0557144	B0557144 AGENCOURT
18	1442.5	10.9	984	5	B0690869	B0690869 AGENCOURT
19	1440.5	10.9	1159	4	BM802749	BM802749 AGENCOURT
20	1420	10.7	1022	5	BM910785	BM910785 AGENCOURT
21	1382.5	10.5	1030	4	BM558844	BM558844 AGENCOURT
22	1378.5	10.4	1010	5	BM915686	BM915686 AGENCOURT
23	1374	10.4	887	5	B0542258	B0542258 AGENCOURT
24	1368	10.4	875	7	CK769773	CK769773 957873 MA
25	1366.5	10.3	868	5	BM472005	BM472005 AGENCOURT
26	1366.5	10.3	868	5	BM472005	BM472005 AGENCOURT
27	1348.5	10.2	923	5	B0184403	B0184403 AGENCOURT
28	1346.5	10.2	1038	5	BM910704	BM910704 AGENCOURT
29	1338	10.1	887	6	CA979881	CA979881 AGENCOURT
30	1334	10.1	923	5	BX390462	BX390462 BX390462
31	1332	10.1	875	5	BQ711119	BQ711119 AGENCOURT
32	1331.5	10.1	842	5	B0224569	B0224569 603400211
33	1326	10.0	1094	4	BM560912	BM560912 AGENCOURT
34	1325.5	10.0	905	5	BQ897825	BQ897825 AGENCOURT
35	1321	10.0	882	5	B0180236	B0180236 AGENCOURT
36	1305.5	9.9	898	5	B0172348	B0172348 AGENCOURT
37	1305	9.9	1045	4	BM560255	BM560255 AGENCOURT
38	1297.5	9.8	833	7	CK778015	CK778015 96516 MA
39	1274	9.6	875	5	BQ691710	BQ691710 AGENCOURT
40	1272	9.6	3575	3	BSM802410	BSM802410 Homo sapi
41	1267	9.6	769	5	BX368972	BX368972 BX368972
42	1262.5	9.6	915	7	CF272442	CF272442 AGENCOURT
43	1260	9.5	1075	4	BM451469	BM451469 AGENCOURT
44	1259.5	9.5	785	7	CO896175	CO896175 Boven 24
45	1256.5	9.5	928	5	B0164114	B0164114 AGENCOURT

ALIGNMENTS

RESULT 1	AY412686	7372 bp	DNA	linear	GSS 16-DEC-2003
LOCUS	AY412686				
DEFINITION	Homo sapiens NCOR2 gene, VIRTUAL TRANSCRIPT, partial sequence,				
ACCESSION	AY412686				
VERSION	AY412686.1	GI:39768651			
KEYWORDS	GSS.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sinsky,J.J., Adams,M.D. and Cargill,M. (2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA				
TITLE	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA				
JOURNAL	Science 302 (5652), 1960-1963 (2003)				
PUBMED	14671302				
REFERENCE	2 (bases 1 to 7372)				
AUTHORS	Clark,A.G., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sinsky,J.J., Adams,M.D. and Cargill,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA				
COMMENT	This sequence was made by sequencing genomic exons and ordering them based on alignment.				
FEATURES	Location/Qualifiers				
source	1..7372				

gene
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>7372
/gene="NCOR2"
/locus_tag="HCM4640"

ORIGIN

Alignment Scores:

Pred. No.:	0	Length:	7372
Score:	8326.00	Matches:	1708
Percent Similarity:	69.56%	Conservative:	1
Best Local Similarity:	69.52%	Mismatches:	748
Query Match:	62.25%	Indels:	0
DB:	9	Gaps:	0

US-09-522-753-5 (1-2517) x AY12686 (1-7372)

QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTPArgAlaThrGluProArgTyrPro 20
DB 1 ATGTGGGATTCACACAGCCTGTGCACAGACGTGGAGGGCCACTAGAGCCCGCTACCG 60
QY 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
DB 61 CCCCAAGCCTTCTTCTTACCCAGTGCAGATGCCGAGACGACACGACGCGGCTCTG 120
QY 41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
DB 121 GAGTACACACACACCTCCCGACATATGCTCCCACTGTCCGCGGCTCCATCATCCAG 180
QY 61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyValangluArgSerGln 80
DB 181 CCCCAAGCCTGAGGAGGCTCCCTGCTGTGAGTTCCAGCCGCGAGATGAACGNNNNNNN 240
QY 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGluLysSerGluMet 100
DB 241 NNN 300
QY 101 GluPheIleGlySerLysArgProArgLeuGluLeuLeuProArgProLeuLeuArgPro 120
DB 301 NNN 360
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
DB 361 NNN 420
QY 141 LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
DB 421 CTGACCGGCAAGCTGGACCGGTGTCTCCCGCCAGCCCGCCGACACTGACCTTGAGCTG 480
QY 161 GluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAsp 180
DB 481 GAGCTGTGCTCCCGCCAGCTGTCCAAAGAGAGACTATCCAGAACATGACCGCGTGGAC 540
QY 181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeu 200
DB 541 CCAAGAGATACCATGTGTAGAGCAGCAGATCTTAAGCTGAAGAAAGACAGCAACAGCTG 600
QY 201 GluGluGluAlaAlaLysProProGluProGluLysProValSerProProIleGln 220
DB 601 GAGAGAGAGAGCTGCCAAGCGCGCGAGCTGAAGAACCCCGTGTACCGCGCCCATCGAG 660
QY 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGlnAla 240
DB 661 TCGAAGCACCGCAGCTGTGTGCAGATCATCTACGAGAGAACCGGNNNNNNNNNNNNNN 720
QY 241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrArgGlnPro 260
DB 721 NNN 780
QY 261 SerAspThrArgGlnTyrHisGluAsnIleLysGlnIleAsnGlnAlaMetArgLysLysLeu 280
DB 781 TCCGACACCCGGCAGTATCATGAGAAATCAAAATAAACAGGCGATGCGGAAGAGCTA 840

QY 281 IleLeuTyrPheLysArgArgHisHisAlaArgLysGlnTrpLysGluLysPheCysGln 300
DB 841 ATCTTGTAATCTCAAGAGAGAAATCACGCTCGAAACAATGGAGACAGAAATTCTGCCAG 900
QY 301 ArgTyrAspGlnLeuMetGlnAlaLeuGluLysLysValGluAlaGlnIleGluAsnAspPro 320
DB 901 CGCTATGACACACTCATGAGAGGCTGTGGAGAAAGAGTGGACGCTATCCAGAACACCCC 960
QY 321 ArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340
DB 961 CGCGCGCGCGCCAAAGAGAGCAAGTGGCGGAGTACTAGAGAGAGTCCCTGAGATC 1020
QY 341 ArgLysGlnArgGluLeuGlnGlnLysArgMetGlnSerArgValGlyGlnArgLysSerGly 360
DB 1021 CCAAGCAGCGCAGAGCTGCAGAGGCATGACAGACAGCGGTGGCCAGCGGCGAGTGGG 1080
QY 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
DB 1081 CTGTCCATGTCCGCGCGCCGACGACGAGAGTGTCAAGATCATCGATGAGCTCTCA 1140
QY 381 GluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400
DB 1141 GAGCAGAGAACTTGGAGAGACAGATGCCCAAGCTGTGCCGTATCCCGCCATGCTGTAC 1200
QY 401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420
DB 1201 GACGCTGACACACAGGCGCATCAAGTTCAATCAATGAAGGCGCTTATGGCCGACCCCATG 1260
QY 421 LysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluTrpPhe 440
DB 1261 AAGGTGTACAAAGACCGCAGGTCATGAACATGTGAGAGACAGAGAAAGAACCTTC 1320
QY 441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
DB 1321 CCGGAGAAANN 1380
QY 461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480
DB 1381 NNACAGTGTGCTGAGTGCCTCTTACTTACTGACTGAAGAGATGAGAACTATAG 1440
QY 481 SerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGln 500
DB 1441 AGCCTGTGTGAGACGAGAGCTATGCGCGCGCGCAAGAGCAAGNNNNNNNNNNNNNN 1500
QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
DB 1501 NNN 1560
QY 521 AspGluLysGluLysGluAlaGluLysGluGluGluLysProGluValGluAsn 540
DB 1561 NNN 1620
QY 541 AspLysGluAspLeuLeuLysGluLysThrAspArgThrSerGlyGluAspAsnAspGlu 560
DB 1621 NNN 1680
QY 561 LysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGly 580
DB 1681 AAGAGGCTGTGGCTCCAAAGCCCGCAAAATGCGCAACAGCAAGGAGAGAGCAAAAGG 1740
QY 581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600
DB 1741 CCATCAACCCGCTCATATGAGGCAACAGAGAGAGGATTCACCCCGCCAGCAG 1800
QY 601 SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGlnGluGluMet 620
DB 1801 AGCGCGAGCTGCTCCATGAGCTGATGAGAGTTCGCTGGAACAAGAGAAAG 1860
QY 621 GluThrAlaLysLysGluLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMet 640
DB 1861 GAAACAGCCAAAGAAAGTCTCTTGAAACAGCGCGCGCACTGCTGCGCATTCGCGGAGT 1920

QY 641 ValGlySerIysThrValSerGlnCysIysAsnPheTyrPheAsnTyrLysValArgGln 660
Db 1921 GTGGGCTCCCAAGACTGTGTGCGACTGTAAAGAACTTCTACTTCACTCAAGAAAGAGCGAG 1980
QY 661 AsnLeuAspGluIleLeuGlnGlnHisIleValLeuIysMetGluIysGluValArgAsnAlaArg 680
Db 1981 AACTCCATGATGATCTTGACAGACACAAAGCTGAAGAGAGAGAGAGAGAGAGAGCGCGG 2040
QY 681 ArgIysIysIysIysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValVal 700
Db 2041 AGGAAAGAAAGAAAGAGCGG 2100
QY 701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGlu 720
Db 2101 GAGGATGAGAGAGAGAGAGCGCTCGCGCGTGAAGCGAATGAGAGAGAGAGAGAGAGAG 2160
QY 721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGluGluCysSerGlyProAla 740
Db 2161 GCTGAAGNN 2220
QY 741 ThrValAsnAsnSerAspThrGluSerIleProSerProHisThrGluAlaAlaLys 760
Db 2221 NNN 2280
QY 761 AspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProPro 780
Db 2281 NNN 2340
QY 781 GlyProProThrProProAlaArgThrSerArgAlaProIleGluProThrProAlaSer 800
Db 2341 NNN 2400
QY 801 GluAlaThrGlyAlaProThrProProProAlaProProSerProSerAlaProProPro 820
Db 2401 NNN 2460
QY 821 ValValProLysGluGluIysGluGluGluThrAlaAlaAlaProProValGluGluGly 840
Db 2461 NNN 2520
QY 841 GluGluGluIlyProProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGlu 860
Db 2521 GAGAGAGAGAGAGCGCGCGCGCGCGCGAGAGCTGGCACTGACACAGAGGAGCGCGAGAG 2580
QY 861 ProValIysSerGluCysThrGluGluAlaGluGluGluProAlaLysGluLysAspAla 880
Db 2581 CCGCTCAAGAGCGAGTGCAGAGAGAGCGCGAGAGGCGCGCGCAAGGCGAGAGCGG 2640
QY 881 GluAlaAlaGluAlaThrAlaGluGluGlyAlaLeuLysAlaGluLysGluGluGlySer 900
Db 2641 GAGGCGGCTGAGGCGACGCGCGAGGGGGCGCTCAAGGCGAGAGAGAGAGGGGCGGAGC 2700
QY 901 GlyAlaArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThr 920
Db 2701 GCGAGGCGCCACACAGCGCAAGAGCTCGGGCGCGCGCGCGCGCGAGCAGCGACTCAGTCA 2760
QY 921 CysSerAlaAspGluValAspGluValaGluGlyGlyValAspLysAsnArgLeuLeuSerPro 940
Db 2761 TCGAGTGCAGACAGAGTGAATGAGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCA 2820
QY 941 ArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysPro 960
Db 2821 AGGCCCGAGCTCTCAACCGCGACTGGCGACCGCGCGGCGCAATGCTCTCAACCGCGAGCCA 2880
QY 961 LeuAspLeuLysGluLeuLysGluGlnArgAlaAlaAlaIleProProIleGlnValThrLys 980
Db 2881 CTGGAGCTGAAGAGCTGAAGAGCGAGCGAGCGGCTGCCATCCCGCGCAACCGGTCAACAAA 2940
QY 981 ValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaProPro 1000
Db 2941 GTCCATGAGCGCGCGCGAGAGAGCGAGCTCCACCAAGCAGAGCTCCCGCGCGCGCGCG 3000
QY 1001 ProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProArg 1020

Db 3001 CCACCGCAAAAACCTGCAGCGCGAGAGAGCGCGCGCTCGAGCGCTGCAGAGAGCGCGCG 3060
QY 1021 GlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGln 1040
Db 3061 GGCAGAGAGAGAGAGCG 3120
QY 1041 LysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValProProArg 1060
Db 3121 NNN 3180
QY 1061 GluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaProPro 1080
Db 3181 NNN 3240
QY 1081 GlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProPro 1100
Db 3241 NGTCACCCACTGCCCCCTGGGCTTCATGACATGGCCCGCGCGCTGCGCGCGCGCGCC 3300
QY 1101 ThrIleSerAsnProProProLeuIleSerSerAlaLysHisAspSerValLeuGluArg 1120
Db 3301 ACCATCTCCAACCGCGCTCCCTCATCTCTCTGCGCAAGCACCCCAAGCTCTCGAAGG 3360
QY 1121 GlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGluHis 1140
Db 3361 CAATAGTGCCATCTCCCAAGAGATGTCGATCCAGCTCCACGTCGCCGTACTCAGAGCAT 3420
QY 1141 AlaValAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLys 1160
Db 3421 GCCAAGGCGCGCGGCGCTGTCAACATGAGGCTGCGCTGCGCGCGCGCGCGCGCGCG 3480
QY 1161 LeuAlaProPheSerGlyValLysGlnGluGluLeuSerProArgGlyGlnAlaGlyPro 1180
Db 3481 CTGGCACTCTTCAAGCGAGAGTGAAGCAGAGCAGCTGTCCCAAGGCGCGCGCGCGCGCA 3540
QY 1181 ProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeu 1200
Db 3541 CCGAGAGAGCTGGGGGGGCCACAGCGCGCAGAGGCGTCCGTGTGAGAGGAGAGCGCTG 3600
QY 1201 GlySerValProGlyGlySerIleThrLysGlyTyrProSerThrArgValProSerAsp 1220
Db 3601 GGCTCAGTTCGGGCGGAGACATCACCAAGGAGATTCGCCAGCACCGGCGCTCGGAC 3660
QY 1221 SerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLys 1240
Db 3661 AGCGGCATCATACCGCGCTCATCACCGCGCACCGCGCAGCTGACGCTCTGTACAG 3720
QY 1241 GlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGlu 1260
Db 3721 GGCACCATCACCAAGATCATCGCGAGGAGCAGCCGAGTGCCTTGGACCGCGCGCGGAG 3780
QY 1261 AspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyr 1280
Db 3781 GACAGCGCTGCCAAGGCGCACGTCATCTAAGAGGAGAGAGGCGCACGCTTGTCTTAT 3840
QY 1281 GluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyLysSerSerSerGlyPro 1300
Db 3841 GAGGNN 3900
QY 1301 ProHisGluThrAlaAlaProLysValArgThrTyrAspMetMetGluGlyArgValGlyArg 1320
Db 3901 NNN 3960
QY 1321 AlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHis 1340
Db 3961 NNN 4020
QY 1341 SerProHisIleLeuLysGluGlnHisIleArgGlySerIleThrGlnGlyIlePro 1360
Db 4021 AGCGCCCAACCACTTCAAGAGAGAGCACCATCTCGCGGCTCTTACACAGAGNNNNNNNN 4080
QY 1361 ArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLysArg 1380

[illegible][illegible]

Qy	2101	AlahsleuProhlaAleuArgProleuProgluSerGlnProSerSerSerProleuLeu	2120
Db	6301	GCCCACTCCACACACTCGCGCGCTGCTGAGACCGACCTGTCACAGCCCGCTCTC	6360
Qy	2121	GlnThrAlaProGluValIleGluYhiIsglnArgValValThrLeuAlaGlnhiIrlIeSer	2140
Db	6361	CAGACCGCGCCAGGGGCTCAAGGTCACCAAGCGGGTGGTCACTTGCCGACGACATCAGT	6420
Qy	2141	GluValIleThrGlnAspTyrThrArgHisIleSProGlnGlnLeuSerAlaProLeuPro	2160
Db	6421	NN	6480
Qy	2161	AlaProLeuTyrSerPheProGluYlaSerCysProValLeuAspLeuArgArgProPro	2180
Db	6481	NN	6540
Qy	2181	SerAspLeuTyrLeuProProProAspHisGluYalaProAlaArgGlySerProHisSer	2200
Db	6541	NN	6600
Qy	2201	GluGlyGlyIlybArgSerProGluProAsnIlyThrSerValLeuGlyGlyIlyGluAsp	2220
Db	6601	NN	6660
Qy	2221	GlyIlyleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAla	2240
Db	6661	GGTATTTGAACCTGTGTCCCAACCGAGGGCATGACGGGCGCAGCTCCCGAGTGTCT	6720
Qy	2241	ValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerIlys	2260
Db	6721	GTTGTAACCGCTGCTGTACCGGAGTGGGAGACAGAGGAGCCAGCAGGATGGCTCCAG	6780
Qy	2261	SerProGlyAsnThrSerSerGlnProProAlaPhePheSerIlySleuthGluSerAsnSer	2280
Db	6781	TCTCAGGCAACACACAGCCAGCCGACGCTTTCTTCAAGACGTCGACCGAGCACTCC	6840
Qy	2281	AlaMetValIlybSerIlybGlyGlnIlyleAsnIlybIlyLeuAsnThrHisAsnArgAsn	2300
Db	6841	GGCATGTGGCAAGTCCAGAMACAGAGATCAACAAGATCGAACCACACACCGGAT	6900
Qy	2301	GluProGluTyrAsnIleSerGlnProIlyThrGluIlePheAsnMetProAlaIleThr	2320
Db	6901	GAGCCTGAAATTCANN	6960
Qy	2321	GlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMet	2340
Db	6961	NN	7020
Qy	2341	GlyLeuGluAlaIleIleArgValAleuMetGlyIlyTyrAspGlnTyrGluGluSer	2360
Db	7021	NN	7080
Qy	2361	ProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAla	2380
Db	7081	NN	7140
Qy	2381	MetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGly	2400
Db	7141	NN	7200
Qy	2401	GlyIlybAlaIyValSerGlyArgProSerSerArgIlybAlaIySerProAlaProGly	2420
Db	7201	GGGAAGGCAAGTCTGTGGCAGACCCAGCAGCCGAAAGCCAGTCCCGGCCCGGGC	7260
Qy	2421	LeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluGlyAspCysAsn	2440
Db	7261	CTGCATCTGGGGAGCCGCGCACCTGTCTCTCTCAGGCACTCGAGGGAGACTGCAAC	7320
Qy	2441	ArgArgThrProLeuThrAsnArgValTyrGluAspArgProSerSerAla	2457
Db	7321	CGCGGAGCGCGCTCAACCAACCGGTGTGGAGAGACAGGCCCTGTCTCGCA	7371

LOCUS	AY412688	7013 bp	DNA	linear	GSS 16-Dec-2000
DEFINITION	Mus musculus NCOR2 gene, VIRTUAL TRANSCRIPT, partial sequence,				
ACCESSION	AY412688				
VERSION	AY412688.1	GI:39768653			
KEYWORDS	GSS.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 7013)				
	Clark,A.G., Gianowski,S., Nielson,R., Thomas,P., Kejaritwal,A.,				
	Todd,M.A., Tarenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B.,				
	Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,				
	Adams,M.D. and Cargill,M.				
	inferring nonneutral evolution from human-chimp-mouse orthologous				
	gene trios				
TITLE	Science 302 (5652), 1960-1963 (2003)				
JOURNAL	14671302				
PUBMED	2 (bases 1 to 7013)				
REFERENCE	Clark,A.G., Gianowski,S., Nielson,R., Thomas,P., Kejaritwal,A.,				
AUTHORS	Todd,M.A., Tarenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B.,				
	Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,				
	Adams,M.D. and Cargill,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,				
	Rockville, MD 20850, USA				
COMMENT	This sequence was made by sequencing genomic exons and ordering				
	them based on alignment				
FEATURES	Location/Qualifiers				
source	1..7013				
	/organism="Mus musculus"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:10090"				
gene	<1..>7013				
	/gene="NCOR2"				
	/locus_tag="HCM4640"				
ORIGIN					
Alignment Scores:					
Pred. NO.:	0	Length:	7013		
Score:	7109.50	Matches:	1509		
Percent Similarity:	63.77%	Conservative:	59		
Best Local Similarity:	61.37%	Mismatches:	768		
Query Match:	53.80%	Indels:	125		
DB:	9	Gaps:	17		
US-09-522-753-5 (1-2517) x AY412688 (1-7013)					
QY	1 MetSerGlySerThrGlnLeuValAlaGlnThrTPARgAlAThrGluProArgTyrPro	20			
DB	1 ATGTCAGAGATCCACACAGCCTGTGGACACAGACATGGCGGCTGCTGAGCCCGCTACCCA	60			
QY	21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu	40			
DB	61 CCCCATGCGATCTCTCAACCCGGTGAGATAGCCCGGCTCCACACGGACGTGGGGCTGCTT	120			
QY	41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln	60			
DB	121 GAGTACCAACACACACCCCGTGACATCACCTGCACACCCGGTTCATCATCCAG	180			
QY	61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln	80			
DB	181 CCACGAGAGAGCGGCCCTTCACTGCTGTCAAGATTCCAGCCTGGGAGTAAAGNNNNNN	240			
QY	81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet	100			
DB	241 NNN	299			
QY	101 GluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro	120			
DB	300 NNN	359			

[illegible]

QY	481	SeRleuValAlrGaGseRerYrAaGaRrAaRgLyLysSeRrGInGInGInGInGInGIn	500
Db	1440	AGCTTGGTGAAGCGCGAaCTATCGCGCGCGTGGCAAGCAaGNNNNNNNNNNNNNNNN	1499
QY	501	GInGInGInGInGInGInGInGInGInGInProMeCProArGseRerGInGInGInGIn	520
Db	1500	NN	1559
QY	521	AspGInuYbGInuYbGInuYbGInuAluYbGInuGInuGInuYbProGInuAlGluAsn	540
Db	1560	NN	1619
QY	541	AspLySGInuAspLeuLeuLySGInuYbThrAspAspThrSeRrGInuAspAsnAspGIn	560
Db	1620	NN	1679
QY	561	LySGInuAlaValAlaSeRrLySGInuYbArGInuYbThrAlaAsnSeRrGInuYbAlaGInuYbGIn	580
Db	1660	AAAGAGGCGGCGCTCCAAAGGCGGCAAACTGCGCAACAGCCAAAGCCCGCGCAAGGC	1739
QY	581	ArGInuThrArGseRerAlaAsnGInuAlaAsnSeRrGInuAlaIleThrProGInGIn	600
Db	1740	CGTATCAaGCGCTCCATaGGCCAAAGAGGCCAACATGaGAGAGACAGCCACCCACAGCAA	1799
QY	601	SeRrAlaGInuAlaSeRMeTAlaSeRMeTGluleuAsnGInuSeRerArGrThrThrGInuGInuMeT	620
Db	1800	AGTTCAAGAGCTGCGCTTCATGaGATGAAGAGAGTTCGCTGCGATGAGAGAAATG	1859
QY	621	GInuThrAlaLySGInuYbGInuLeuGInuHISGInuYbArGInuThrSeRrAlaIleAlaArGMeT	640
Db	1860	GAGACAGCAAAAGAAAGGCTCTCTGAACATGGAGAGAACTGGTCAGCCATTGCCCGCATG	1919
QY	641	ValGInuSeRrLyThrValSeRrGInuCyLeuAsnPhenYrPhenAsnYrLySGInuYbArGIn	660
Db	1920	GTGGGCTCCAGAGCCGTGTCCCAAGTGAAGACTTCACTTCAACTCAACTCAAGAGAGGCAG	1979
QY	661	AsnLeuAspGInuIleLeuGInGInHISLySGInuYbMeTGluleuGInuAlaGInuAlaArG	680
Db	1980	AACTGAGCGAAATCTTTCAGACACACAAGCTTAAAGATGAGAGAGAGAGAAAGCGCTCGG	2039
QY	681	ArGInuYbLySGInuYbAlaProAlaAlaSeRrGInuAlaAlaPheProProVal	700
Db	2040	AGGAAGAAAGAAAGAACCCCGAGCGCGGCGAGGAGAGACAGGCTTCCACCTGCCGCT	2099
QY	701	GInuAspGInuMeTGluleuAlaSeRrGInuValSeRrGInuYbAsnGInuGInuMeTValGInu	720
Db	2100	GAGGACGAAGAGATGGAAGCATCAGCGCGCAAGTCCCAATGAGAAAGAGCTGGCGGAGAG	2159
QY	721	AlaGInuAlaLeuHISAlaSeRrGInuYbAsnGInuValProArGInuGInuYbSeRrGInuProAla	740
Db	2160	CGAGAGAGNN	2189
QY	741	ThrValaAsnAsnSeRerAspThrGInuSeRrIleProSeRrProHISThrGInuAlaAlaLyS	760
Db	2190	NN	2249
QY	761	AspThrGInuYbAsnGInuYbProLySGInuYbProAlaThrLeuGInuAlaAspGInuProPro	780
Db	2250	GAG-----	2252
QY	781	GInuProProThrProProArGInuYbThrSeRrAlaProIleGInuProThrProAlaSeR	800
Db	2252	-----	2252
QY	801	GInuAlaThrGInuAlaProThrProProProAlaProProSeRrProSeRrAlaProProPro	820
Db	2252	-----	2252
QY	821	ValValProLySGInuGInuYbGInuGInuThrAlaAlaAlaProProValGInuGInuY	840
Db	2253	-----CAGAAAGTCTGAGGCGGAGAG-----ATCGATGTGGCA	2285
QY	841	GInuGInuGInuYbProProAlaAlaGInuGInuAlaValAspThrGInuYbAlaGInuGInu	860

Db 2286 AACCCAGAGGAGCCGAGGCTCTGAGAG-----CCCCCGAG 2324
Qy 861 ProValIysSerGlySerThrGluGluGluGluProAla---LysGlyLysAsp 879
Db 2325 AGCTTAAAGGTGACCAAGAGAGACCGAGAAAGCTTGAAAGCAAGCCAGGCGC 2384
Qy 880 AlaGluAlaIaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysGluGly 899
Db 2385 ACAGAGGCCATTGAATGTGTCTGAGGCACCACTTAAGGTGAG-----GAGGCTGCT 2438
Qy 900 SerGlyArgAlaThrThr---AlaLysSerSerGlyAlaProGluAspSerAspSer 918
Db 2439 AGCAAGGCAAGCTGTGACCAAGGGTTCCAGCTCAGTCCACCCAGACAGACATGACTCCAGT 2498
Qy 919 AlaThrCysSerAlaAspGluValaAspGluAlaGluGlyGlyLysAspLysAlaGluLeu 938
Db 2499 GCCACCTGAGTGCCTGATGAGTGAAGCAACCGAAGAGAGTGAAGAGGCGAGCTGTCTG 2558
Qy 939 SerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAspAlaSerProGln 958
Db 2559 TCACCAAGGCCCAAGCTCTCCACCCGGCTGAGAGATCCCGGGCCAGTACTGCCCCAG 2618
Qy 959 LysProLeuAspLeuLysGluLeuLysGluArgAlaAlaAlaIleProProlIleGlnVal 978
Db 2619 AAGCCCTGGACTGAAGCAGTGAAGCAGCAGCAGCCGCAATCCCCCTATCCAGGTC 2678
Qy 979 ThrIleValAlaIleGluProProArgGluAspAlaAlaProThrLysProAlaProProlAla 998
Db 2679 ACCAAGGCTCATAGCCCCCGGAGAGACACAGTACCCCAAGCAAGTTCCTCCCTGTG 2738
Qy 999 ProProProProGluAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSer 1018
Db 2739 CCTCCACCCACGACGACCTACAGCCAGAGGAGTGACTGTCTCAGCAGTGGAGAGAGT 2798
Qy 1019 ProArgGlyLysSerArgSerProAlaProProlAlaAspLysGluAlaAlaGlu 1038
Db 2799 CCACGTGGCAAGTCCCGCAGCCAGCTGCTCCGCAAGAAAGGNNNNNNNNNNNNNN 2858
Qy 1039 AlaGluLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValPro 1058
Db 2859 NNN 2918
Qy 1059 ProArgGluValIleLysAlaSerProAlaAlaProAspProSerAlaPheSerTyAla 1078
Db 2919 NNN 2978
Qy 1079 ProProGlyAlaIleProLeuProLeuGlyLeuAlaAspThrAlaArgProValLeuProArg 1098
Db 2979 NNN 3038
Qy 1099 ProProThrIleSerAsnProProProlLeuIleSerSerAlaLysAspProSerValLeu 1118
Db 3039 CCCCCC---ATCTCTAAACCCCAACCCCTCATCTCTGCAAGACATCCCGCGTACTT 3095
Qy 1119 GluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuAlaValProTySer 1138
Db 3096 GAGAGGCACTGGGTGCTCATCTCCAGGGAGATGTCAGTCCAGCTTGCTGCTTCACTCA 3155
Qy 1139 GluAlaIleLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspPro 1158
Db 3156 GAGCAAGCCAGGCCCCCATGGGCTCTCACCATGGGGCTGCCCTTGGCGTGAACCT 3215
Qy 1159 LysLysLeuAlaAspPheSerGlyValLysGlnGlnLeuSerProArgGlyGlnAla 1178
Db 3216 AAGAACTGAGCCCTTTCAGTGAAGTGAAGCAAGAAAGTTGTCTCCCTCGGGTTCAGGCT 3275
Qy 1179 GlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuAspGlyThr 1198
Db 3276 GGGCCACCTGAAGTCTGGAGAGTCCACATGCTCAGAGAGACCTGTGCTGAGAGGGACA 3335
Qy 1199 AlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValPro 1218

Db 3336 GCACTGGGCTCCGCCACCACTGGAAGCATCACCAAGGAGCTCCCACTACCGGGCTGCA 3395
Qy 1219 SerAspSerAlaIleThrTyArgGlySerIleThrAlaGlyThrProAlaAspValLeu 1238
Db 3396 GAGGAGCCCC-----ACCTACAGAGGCTTATCACCCAGCGACGCGCGAGAGCTCTC 3449
Qy 1239 TyrLysGlyThrIleThrArgGlyIleGlyLysAspSerProSerArgLeuAspArgGly 1258
Db 3450 TACAGAGGTACATCAGCAGAGATGTCGTGAAGACAGCCCAAGTCCCTTGACCGGACA 3509
Qy 1259 ArgGluAspSerLeuProLysGlyLysValIleTyTrpGluGlyLysLysValLeu 1278
Db 3510 CGAAGAGACACCTGCGCCCAAGGCGCATGTCTATTAAGGCAAGAAAGGCACAGTCTTA 3599
Qy 1279 SerTyTrpGluGlyLysMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSer 1298
Db 3570 TCCTATGAAGNN 3629
Qy 1299 GlyProProAlaIleGluThrAlaAlaProLysArgThrTyAspMetGluGlyArgVal 1318
Db 3630 NNN 3689
Qy 1319 GlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGlu 1338
Db 3690 NNN 3746
Qy 1339 ArgHisSerProHisLysLeuLysGluGlnHisAlaIleArgLysSerIleThrGlnGly 1358
Db 3747 CAGACAGCCCC---CACTCAGAGAGACATCATCCGAGGCTCATCAGCCAAAGN 3803
Qy 1359 IleProArgSerTyValGluAlaGlnGluAspTyLeuArgArgGluAlaLysLeuLeu 1378
Db 3804 NNNNN---NN 3862
Qy 1379 LysArgGluGlyThrProProProProProProSerArgAspLeuThrGluAlaTyLys 1398
Db 3863 NNNNNN-----NN 3895
Qy 1399 ThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrVal 1418
Db 3896 NNN 3955
Qy 1419 LysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGlu 1438
Db 3956 NNN 4015
Qy 1439 LeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLys 1458
Db 4016 NNN 4075
Qy 1459 TyrAspThrGlyAlaSerThrThrArgLysTyLysHisAspValArgSerLeuIleGly 1478
Db 4076 NNN 4135
Qy 1479 SerProGlyArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAla 1498
Db 4136 AGCCCGGCGGCGCTTTCCTGCGCGACACCGCTGAGACATATGCTGATGCCGGGCA 4195
Qy 1499 LeuGluArgAlaCysTyTrpGluGluSerLeuLysSerArgProGlyThrAlaSerSerSer 1518
Db 4196 CTGGAGCGTGGCTCTATGAAGAGCTGAAGAGCCGGTCAAGAGACACAGACTGGTGCA 4255
Qy 1519 GlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgGln 1538
Db 4256 GGGGGCTTCATCACAGTGGGCTTCCAGTGTGTGTCTGAACTGGGCAAGCCAGGCAA 4315
Qy 1539 SerProLeuThrTyGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySer 1558
Db 4316 AGCCCACTGACTTACGAAGACACGCGGACCTTTCACCAAGTCACTGCACTGCTCC 4375
Qy 1559 ProValThrMetCysArgGluProThrProArgLeuGlnGluGlySerLeuSerSerLys 1578
Db 4376 CTTGTGACACGAGGAGGCCACGCGCTTTCAGGAAAGNNNNNNNNNNNNNNNNNNNNNN 4435

QY 1579 AlaSerGlnAspArglyLeuThrSerThrProArgGluIleAlaIysSerProHisSer 1598
 DB 4436 NNN 4495
 QY 1599 ThrValProGluHisIleProHisProIleSerProTyGluHisIleuLeuArglyVal 1618
 DB 4496 NNN 4555
 QY 1619 SerGlyValAspLeuTyArgSerHisIleProLeuAlaPheAspProThrSerIlePro 1638
 DB 4556 NNN 4615
 QY 1639 ArgGlyIleProLeuAspAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1658
 DB 4616 NNN 4675
 QY 1659 ProThrTyProHisIleuTyProProTyTLeuIleArgGlyTyProAspThrAlaAla 1678
 DB 4676 CCACCTTACCCACACCTTACCCACCTTACCTTACCTTACCTTACCTTACCTTACCTTAC 4735
 QY 1679 LeuGluAsnArgGlnThrIleIleAsnAspTyTleThrSerGlnGlnMetHisAsn 1698
 DB 4736 CTGGAGAACCGGCAGACCATCATATGACTACATCACTCGAGAGATGACCAAC 4795
 QY 1699 ThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSer 1718
 DB 4796 GGTGCTCCGCCATGCGCCAGCGCTGTGACATGCTAGGGGTGTGTACCGGAGAGTCTC 4855
 QY 1719 SerLeuAlaLeuAsnTyTAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1738
 DB 4856 TGGCTGGCCCTCAATATATCCGCTGGCCCAAGAGCATTATCGACTGTCTCCCAAGTCCA 4915
 QY 1739 HisLeuProValLeuValProProThrProGlyTThrProAlaThrAlaMetAspArgLeu 1758
 DB 4916 CACCTGCGCGGTGTGTGTGCAACCAAGCAGCAGCCCTGCAACCCCATCGACCGCTT 4975
 QY 1759 AlaTyLeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSer 1778
 DB 4976 GCTACTCCGCCATGCGCCAGCCCTTACAGAGCGGCGCACATGACTCACCGCTGTCTC 5035
 QY 1779 ProGlyGlyProThrHisIleuThrTyProThrTThrSerSerSerGlyuArgGlyuArg 1798
 DB 5036 CCAAGAGGCCCACTCACTAGCTAAACCAACTGACCACTTTCATCGAGCGGGAACCG 5095
 QY 1799 AspArgAspArgGluArgAspArgAspArgGlyuArgGlyuSerIleuThrSerThr 1818
 DB 5096 GAACGTGACGGGAGAGAGC-----AAGTCAATCTCACGTCTACC 5137
 QY 1819 ThrThrValGluHisAlaProIleTArgProGlyTThrGluGlnSerSerGlySerSer 1838
 DB 5138 ACTACAGTGGAGATCACCCCATCTGAGACCTGTACGGAGCAGAGCGGGCT----- 5194
 QY 1839 GlySerSerGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHis 1858
 DB 5195 -----GGGGGCGAGCGCGCGCGCTCCACAC-----CAC 5227
 QY 1859 GlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeu 1878
 DB 5228 CAGCACTGCGCCATCTCCCGGACCCAGAGCGCTTGAGAGAGAGCGCCACATGTGCTG 5287
 QY 1879 HisAsnThrGlyMetIysGlyIleIleThrAlaValGluProSerIysProThrValLeu 1898
 DB 5288 CACAACACAGCATGAAGGCGGTGTCCTCCGTGGAACCGGCAACGCGCAAGTCTCG 5347
 QY 1899 ArgSerThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHis 1918
 DB 5348 AGGTTCACCTCCACCTTGTGCTGTCCGCCAGCTGCACATTTCCACCTCGCACCCAC 5407
 QY 1919 CysProLeuGlyGlyThrLeuAspGlyValTyTProThrLeuMetGluProValLeuLeu 1938
 DB 5408 TGGCCACTTGTGTGACCCCTTGAAGGGGTCTACCTCACTCAAGAGCGCGTCTGTATA 5467

QY 1939 ProTyGluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAla 1958
 DB 5468 CCCAAGAGACCTCTCGGGTCCCGCGCGCGCGCGCGCTGTGGACCTGTGGACCTGTGCC 5527
 QY 1959 PheLeuAlaTyProProAlaArgSerGlyLeuGluProAlaSerSerProSerIysGly 1978
 DB 5528 TTCTTCACCAACCCCGGCGCG-----GAGCGCGCTCTCTCACCCAGCAAGACC 5578
 QY 1979 SerGluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrPro 1998
 DB 5579 TCCGAGCCCCGATTCCTTGACACCCCGCAGCTCAGACACAGCGCATGCGCGCACCCCA 5638
 QY 1999 AlaTyAsnLeuAlaProHisIleAlaSerProAspProProAlaProProAlaSerAla 2018
 DB 5639 GCAAAGAGCTTGACCCACCATGCTCAGTCCGAGCCCGCGCG-----CCCATCTTGGCGC 5695
 QY 2019 SerAspProHisArgGlyuArgThrGlnSerIysProPheSerIleGlnGluLeuGluLeu 2038
 DB 5696 TCGAGTCTGCACCGAAGAAAGACTCAAGTAAACCTTTCCATCCAGAAATTGGAACTC 5755
 QY 2039 ArgSerLeuGlyTyTHisIleGlySerSerTySerProGluGlyValGluProValSerPro 2058
 DB 5756 CGTCTCTCGANN 5815
 QY 2059 ValSerSerProSerLeuThrHisAspTyGlyLeuProTyHisIleuGluGluLeuAsp 2078
 DB 5816 NNN 5875
 QY 2079 TySerHisIleuGluGlyGluLeuArgProTyGlnProGlyProValTyLeuGlyGly 2098
 DB 5876 NNN 5935
 QY 2099 GluAlaAlaHisIleuProHisIleuArgProLeuProGluSerGlnProSerSerSerPro 2118
 DB 5936 GAGCTGCCCATCTCCACATCTGGCGGCATCGCCGAGGCAAGCGCATCTCGACGCCA 5995
 QY 2119 LeuLeuGlnThrAlaProGlyValIysGlyHisGlnArgValValThrLeuAlaGlnHis 2138
 DB 5996 CTCTTCAGACGTGCCCGAGGACTCAAAAGGTCAACAGAGGTGTCTACCTGTGCTCGCAC 6055
 QY 2139 IleSerGluValIleThrGlnAspTyTThrArgHisIleProGlnGlnLeuSerAlaPro 2158
 DB 6056 ATCAGCANN 6115
 QY 2159 LeuProAlaProLeuTyTSerPheProGlyAlaSerCysProValLeuAspLeuArg 2178
 DB 6116 NNN 6175
 QY 2179 ProProSerAspLeuTyTleuProProProAspHisGlyAlaProAlaArgIysSerPro 2198
 DB 6176 NNN 6235
 QY 2199 HisSerGluGlyGlyTyAspArgSerProGluProAsnTyTSerValLeuGlyGlyGly 2218
 DB 6236 NNN 6295
 QY 2219 GluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerSer 2238
 DB 6296 GAGATGCCATTGAGCTGTGTCCCAACAGAGGGATGATCATGAGCAGACATGCTCG 6355
 QY 2239 SerAlaValTyTProLeuLeuTyTArgAspGlyGluGlnThrGluProSerArgMetGly 2258
 DB 6356 ACGCGTGTATCCCATCTGTATCGAGACGGGGAAACAGGGGAGCGCCAGAGGAAGGGC 6415
 QY 2259 SerTySerProGlyAsnThrSerGlnProProAlaPhePheSerTyLeuThrGluSer 2278
 DB 6416 TCTAAGTCTCCAGGCAACCAAGCAGCGCCAGCTTCTTCAAGTACGTGAGAGC 6475
 QY 2279 AsnSerAlaMetValIysSerIysGlnGluIleAsnTyLeuLeuAsnThrHisAsn 2298
 DB 6476 AACTCCGCGCATGTGTAAGTCAAGAGACAGAGATCAACAAGAACTCAACCCCAAC 6535
 QY 2299 ArgAsnGluProGluTyTArgHisIleSerGlnProGlyTThrGluIlePheAsnMetProAla 2318

[illegible]

Pred. No.:	0	Length:	6836
Score:	6533.00	Matches:	1407
Percent Similarity:	61.81%	Conservative:	1
Best Local Similarity:	61.76%	Mismatches:	870
Query Match:	49.44%	Indels:	0
DB:	9	Gaps:	0
US-09-522-753-5 (1-2517) x AY12687 (1-6836)			
Qy	27	ProValGlnIleAlaArgThrIsthArapValGlyLeuLeuGluIuTyGrInhiAhiSer	46
Db	2	CCAGTGAAGATTCCCGGACGACACAGGACGTCGGGCTCTGTGAGTACCAACCACTCC	61
Qy	47	ArgAspTyrAlaSerIsthLeuSerProGlySerIleIleGlnProGlnArgArgPro	66
Db	62	CGGACTATGCTCCCACTGTGCGCGGCTTCATCATCTCAAGCCCAAGCGGAGGCC	121
Qy	67	SerLeuLeuSerGluPheGlnProGlyAengIuAArgSerGlnIuLeuHiALeuArgPro	86
Db	122	TCCCTGCTGTCTAGTTCACGCCCGGAAATGAACGNNNNNNNNNNNNNNNNNNNNNN	181
Qy	87	GluSerHisSerTyrLeuProGluLeuGluIySerGluMetGluPheIleGluSerLys	106
Db	182	NN	241
Qy	107	ArgProArgLeuGluLeuLeuProAspProLeuLeuAArgProSerProLeuLeuAlaThr	126
Db	242	NN	301
Qy	127	GlyGlnProAlaGlySerGluApreLurHysAparaGserLeuThrGlyLysLeuGlu	146
Db	302	NN	361
Qy	147	ProValSerProProSerProProHiIsthAapProGluLeuGluLeuValProProArg	166
Db	362	CCGGTGTCTCCCGCCGACCCCGGACACCACTGAGTGGAGTGTGTCGCGCGCGG	421
Qy	167	LeuSerLysGluGluLeuIleGlnAmetArapValAapAArgGluIleTherVal	186
Db	422	CTGTCCAAAGAGAGCTGATCCAGAAACATGACCGCGTGACCAAGAGATCACCATGTA	481
Qy	187	GluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeuGlnGluGlnAlaAlaLys	206
Db	482	GACACACATCTCTAAGCTGAAGAAAGACGANNNNNNNTGACGAGAGAGCTGCCAG	541
Qy	207	ProProGluProGluLysProValSerProProProIleGluSerLysHiAArgSerLeu	226
Db	542	CCGCCCGAAGCTGAGAAAGCCCGTGTCAACCGCGCCCATCGAGTGAAGACCCGACCTG	601
Qy	227	ValGlnIleIleTyrAapGluAapArgLysAlaGluAlaAlaAhiAArgIleLeuGlu	246
Db	602	GTCGACATCATCTACACACAGAAACCGANNNNNNNNNNNNNNNNNNNNNNNNNNNN	661
Qy	247	GlyLeuGluProGlnIleValGluLeuLeuProLeuTyrAaengIlnProSerAapThrArgGlnTyr	266
Db	662	NNNNNNNNNNNNNNNNNNNNNNNNNNCCGCTGTACAAACAGCCCTCGACACCCGCGAGTAT	721
Qy	267	HisGlnAapIleLysIleAengIlnAlaMetArgLysLysLeuIleLeuTyrPheLysArg	286
Db	722	CATGAAACATCAAAATATAAACAGGCGATGCGGAAAGAACTAAATCTGTACTTCACAGG	781
Qy	287	ArgAapHisAlaAArgLysGlnTyrProGlnLysPheCysGlnAArgTyrAapArgIlnLeuMet	306
Db	782	AGGAATCAAGCTCGAAGACATGGAGGAGAAAGTTCTGCAAGCCCTATGACACAGCTCATG	841
Qy	307	GluAlaLeuGluLysLysValGluAArgIleGluAapAapProAArgAArgAlaLysGlu	326
Db	842	GAGGCTGGGAGAAAGAGGTGAGCATCGAAGAAACACCCCGCGCGGCGCAAGAG	901
Qy	327	SerLysValArgGluTyrTyrGluLysGlnPheProGluIleAArgLysGlnAArgGluLeu	346
Db	902	AGCAAGGTGCGGAGATCTACGAAAGCAAGTTCCTTAGTTCGCAAGCAGCGCGAGCTG	961

OY		347	GInGIuaIgmMeCInSerArYVaIgIyGInaRgISerGIyLeuSerMetSeRlAAla	366
Db		962	CAGGAGCGCATGCACAAGACGGGTGGCCAGCGCGGCGAGTGGTTGTTCATGTCCGCTGCC	1021n
OY		367	AAGSergIUHISGLIVaISerGIuIleIeaSPGYLeuSerGIuNGInUAsnLeuGIu	386
Db		1022	CCGAGCGACACGAAGTGTCAGAGATCATCATGGCCCTTCAGAGACAGAACCCTGGAG	1081n
OY		387	IySGImeArGIuLeuAlaValIlleProPomEtleUTyrASpAlaaSPdngInaRg	406
Db		1082	AAGCAGATGCGCACGTGGCGGTGATCCGGCTTAAGCTGTACAGACGCCGACAGCAGCGC	1141n
OY		407	IlElYSpHeIlAasMeCaSnGIyLeuMeClAaSPrometLySaVAlTyrlYsaSPaRg	426
Db		1142	ATCAAGTTTCATCAAATGAAACGGGCTTATGTAGCTGCACCCCATGAAGTGTAAACAAGACGCG	1201n
OY		427	GInVAlmeASmMeTrSerGIuNGInGIuYsaGIuThrPhaeARGIuLysPheMeCIu	446
Db		1202	CAGGTATGAATACTGGAGTAGAAGAGGAGAAAGACCTTCGGGGAAGANNNNNNNNN	1261n
OY		447	HISPrOlyASaSPheGIyleuIlleaSerPheLeuGIuArGIyThrValaIGluCyS	466
Db		1262	NNNGTGGCTGAGTGC	1321n
OY		467	VAlleUrTYrTYrLeuThrILysSaNGIuaSnTYrLYSeRleuValIRgrSer	486
Db		1322	GTCCCTCTATTACTACCTGACTTAAGAAGATGAGAACCTTAAGAGCGCTGGTAGACGGAGC	1381n
OY		487	TyrATGaRGARGIyLysSerGINgInGINgInGINgInGINgInGINgInGINgIn	506
Db		1382	TATCGGCGCGCGGACAGAGCCAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	1441n
OY		507	GInGINgInGIuIPromEProARgSerSerGINgIuGIuYsaSPGIuLysGIuLysGIu	526
Db		1442	NN	1501n
OY		527	LysGIuAlaGlIuLysGIuGIuLysProGIuValGIuAsnAPlySGIuAspleuEu	546
Db		1502	NN	1561n
OY		547	IySGIuLysThRAspARphrSerGIyGIuAsPaMaSPGIuLysGIuAlaValaIAser	566
Db		1562	NNGGAGAAACAGACGACACCTCAGGGGAGGACACGACGAAGAAGAGCGGTGGCTGCC	1621n
OY		567	IySGIyArGIySThrAlaasnSerGINgIyArGIySLySGIyArgIlleThrARgSerMet	586
Db		1622	AAAAGCCGAAAACTGCCAACAGCCAGGGAAACCGMAAGCCGATCACCCGCTCATG	1681n
OY		587	AlaasngIuAlaasnSerGIuAlaIleThrProGINgsERalAGIuLeuAlaser	606
Db		1682	GCTAATGAGGCCAACAGCAGAGAGGCCATCACCCCCAGACAGCGCCGAGCTGGCTCC	1741n
OY		607	MeGIuLeuAsnGIuseSerARgTrpThrgIuGIuIMetGIuThralayelysgIy	626
Db		1742	ATGGAGCTGAATGAGAGTTCGTGCTGGACAGAAGAAATGGAACACCGCAAGAAAGT	1801n
OY		627	LeuLeuGIuHIISGIyARgAsnrTpsERalaleaIaArGMeValaGISerLythrVal	646
Db		1802	CTCCTGGAAACAGCGCGCACTGGCTGGCCATCGCCCGGATGGTGGCTCCAAAGCTGTG	1861n
OY		647	SerGIuCYLSaSPheTyRpheaSnTYrLysIySArGIuAsnLeuAspGIuIlleLeu	666
Db		1862	TGCGAGTGTAAAGACTTCTACTTCAACTCAAGAGAAGGACAGAACCTCGATGAGATCTTG	1921n
OY		667	GInGINHISLyLeuLysMetGIuLysGIuLysGlaenAlaIRgArGIuLysLylysylala	686
Db		1922	CAGCAGCAACAACCTGAAGATGAGAGAGAGAGAAACGACGAGAGAAAGAAAGAAACG	1981n
OY		687	ProAlaAlaIASerGIuAlaAlaPheProPoValValGIuASPdngIuImetGIu	706
Db		1982	CCGGCGCGCGGACGAGAGGCTGCATTCGCCGCCGTGTGTGAAGATGAGGACATGGAG	2041n
OY		707	AlaSerGIyValSerGIyAsnGIuGIuImetValGIuGIuAlaGIuAlaIeuHIIsAla	726

[illegible]

[illegible][illegible]

```

QY 1807 AaPaTgGluAArgGluLysSerIleLeuThrSerThrThrValGluuHialaProIle 1826
Db 5342 NNNNNNNAGCGGGGAAAAAGGCCATCTCCAGGTCCANNNGTGGAGACGCGCCCATC 5401
QY 1827 TTPaTgProGluYThrGluInserSerGlySerSerGlyGlyGlyGly 1846
Db 5402 TGGAGNNNNNGTACAGAGCAGACGCGGACGAGCGGCGGCGGTGGGGG 5461
QY 1847 SerSerSerArgProAlaSerHisSerHisAlaHisGluHisSerProIleSerProArg 1866
Db 5462 AGAGAGAGCGCGCGCTCCCATCCANNCCACAGACACTCGCCATCTCCCTNNN 5521
QY 1867 ThrGluAspAlaLeuGluInGluArgProSerValLeuHisSerThrGlyMetGlyGlyIle 1886
Db 5522 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5581
QY 1887 IleThrAlaValGluProSerLysProThrValLeuArgSerThrSerThrSerPro 1906
Db 5582 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5641
QY 1907 ValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAsp 1926
Db 5642 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5701
QY 1927 GlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAla 1946
Db 5702 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5761
QY 1947 ArgProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArg 1966
Db 5762 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5821
QY 1967 SerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeuValPro 1986
Db 5822 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5881
QY 1987 ProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHis 2006
Db 5882 CCGGTCTCTGGCCACGCCACCATCGCCCGCACCCCTCGCGAAGAACTCTCGCACCTCACAC 5941
QY 2007 AlaSerProAspProProAlaProProAlaSerAlaSerAspProHisAspArgGluValThr 2026
Db 5942 GCCAGCCCGGACCGCGCGCGGCGGACCTGCTCGGACCTCGGACCGGAAAAAGACT 6001
QY 2027 GluSerLysProPheSerIleGluGluLeuGluLeuArgSerLeuGlyTyrHisGlySer 2046
Db 6002 CAAAGTAAACCTTTTCCATCCAGAACTCGAATCTCGTTCTGGNNNNNNNNNNNNNN 6061
QY 2047 SerTyrSerProGluGlyValGluProValSerProValSerSerProSerLeuThrHis 2066
Db 6062 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6121
QY 2067 AspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGluGluLeu 2086
Db 6122 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6181
QY 2087 ArgProLysGluInProGlyProValLysLeuGlyGlyValAlaHisLeuProHisLeu 2106
Db 6182 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6241
QY 2107 ArgProLeuProGluSerGluInProSerSerSerProLeuLeuGluInThrAlaProGlyVal 2126
Db 6242 CGGCCCTGCTGTAGAGCAGCCCTGCTCGACGCCGCTGCTCCAGACCGCCGAGGAGGTC 6301
QY 2127 LysGluHisGluArgValValThrLeuAlaGluHisIleSerGluValIleThrGluAsp 2146
Db 6302 AAAGGTACACGAGGGGGGTGTACCTGTGGCCNNNNNNNNNNNNNNNNNNNNNNNNNN 6361
QY 2147 TTTThrArgHisHisProGluInGluLeuSerAlaProLeuProAlaProLeuTyrSerPhe 2166
Db 6362 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6421
QY 2167 ProGluAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuPro 2186

```

```

Db 6422 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6481
QY 2187 ProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSer 2206
Db 6482 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6541
QY 2207 ProGluProAspLysThrSerValLeuGlyGlyGlyGlyLysArgGlyIleGluProValSer 2226
Db 6542 CCAGAGCCAAACAGACGTCACTTGGGTGGCGGTGAGAGCGGTATTAACCTGTGTCC 6601
QY 2227 ProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyr 2246
Db 6602 CCACCGAGGAGCATGACGAGGCGGACCTCCGAGTGTGTGTACCCGCTGTATC 6661
QY 2247 ArgAspGlyGluGluInThrGluProSerArgMetGlySerLysSerProGlyAsnThrSer 2266
Db 6662 CCGGATGGGGAACAGACGAGGCCAGGAGATGGGTCTCAATCTCCAGCAACACACAG 6721
QY 2267 GluProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLys 2286
Db 6722 CAGCCGCCAGCTTCTTTCAGCAAGCTGACCGAGACCACTCCGCAATGCTCAAGTCCAG 6781
QY 2287 LysGluGluIleAsnLysLeuAsnThrHisAsnArgAsnGluProGluTyr 2304
Db 6782 AAGCAAGAGATCAACAGAACTGAACACCAACCGAATGAGCTGAAATAC 6835

RESULT 4
BC033087 2000 bp mRNA linear HTC 04-MAR-2003
LOCUS BC033087
DEFINITION Homo sapiens, Similar to nuclear receptor co-repressor 2, clone
IMAGE:5016291, mRNA.
ACCESSION BC033087
VERSION BC033087.1 GI:21619972
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Molecular Function: Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2000)
AUTHORS Strausberg,R.
TITLE Direct Submision
JOURNAL Submitted (25-JUN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contract: MGC help desk
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Boedtel, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Heiao, Martin Krzywinski, Reta Kutche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smalins, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Teal, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: IRRL Plate: 43 Row: i Column: 21
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 5454073
This clone has the following problem: retained intron.
Location/Qualifiers
1..2000
/organism="Homo sapiens"

```

```
/mol_type="mRNA"  
/db_xref="taxon:9606"  
/clone="IMAGE:5016291"  
/tissue_type="lung, small cell carcinoma"  
/clone_id="NH_MGC_7"  
/lab_host="DH10B-R"  
/note="Vector: pOTB7"
```

ORIGIN

Alignment Scores:

Pred. No.:	4,36e-140	Length:	2000
Score:	2566.00	Matches:	503
Percent Similarity:	90.72%	Conservative:	15
Best Local Similarity:	88.09%	Mismatches:	31
Query Match:	19.42%	Indels:	22
DB:	3	Gaps:	1

US-09-522-753-5 (1-2517) x BC033087 (1-2000)

```
Qy 1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro 20  
Db 273 ATGTGGGATTCACACAGCCGTGTGGACAGACGTGGAGGGCCACTGAGCCCCGCTACCG 332  
Qy 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40  
Db 333 CCCACAGCCTTCTCTACCCAGTGCAGATGCGCCGACGACACAGCAGCGAGCTCGGCTCTG 392  
Qy 41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60  
Db 393 GAGTACACAGACCACTCCCGGACTATGCTCCACCTGTCGCGGCTCATCATCAG 452  
Qy 61 ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80  
Db 453 CCCAGGGGGAGGAGCCCTCTCTGCTGTGAGTTCAGCCCGGAGAAAGATGAGTCCAG 512  
Qy 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLeuSerGluMet 100  
Db 513 GAGCTCCACCTGGGGCCAGAGTCCCACTACTGCCCCGAGCGGGGAAAGTCAGATG 572  
Qy 101 GluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120  
Db 573 GAGTTCATTAAAGCAAGCCCTCGGCTAGAGCTGCTGCTGACCCCTGCTCGACCG 632  
Qy 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140  
Db 633 TCACCCCTGTGGCCACGGGCTGCGGATCTAAGACTTCACCAAGACCGTAGC 692  
Qy 141 LeuThrGlyLysLeuGluProValSerProProSerProHisThrAspProGluLeu 160  
Db 693 CTGACGGGCAAGCTGGAACCGGTGTCTCCCCACGCCCGGACACCTGAGCTG 752  
Qy 161 GluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAsp 180  
Db 753 GACTGTGTGCGCCACGCTGTCTCAAGAGAGAGCTGATCCAGAACATGACCGGTGAC 812  
Qy 181 ArgGlnIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeu 200  
Db 813 CGAGAGATTCACCTGTAAGACAGACATCTTAAGCTGAAGAAAGACACACACTG 872  
Qy 201 GluGluGluAlaAlaLysProProGluProGluLysProValSerProProIleGlu 220  
Db 873 GAGAGAGAGCTGCMAAGCCGCCGAGCTGAGAAAGCCCGTCTACGCCGCCATCGAG 932  
Qy 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysValAlaGluAla 240  
Db 933 TCAGAGACACCGACCGCTGTGACAGATCATCAACACAGAACCGGAAGAGCTGAGCT 992  
Qy 241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPro 260  
Db 993 GCACATCGAATTCGGAAGGCTGGGGCCCAAGGTGAGCTGCGCTGTACACACGCCC 1052  
Qy 261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLeu 280
```

```
Db 1053 TCCGACACCCGGCAGATATCATGAAACATCAATAAACAAGGATCGGAAAGACTA 1112  
Qy 281 ILeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGln 300  
Db 1113 ATCTTGACTTCAGAGAGAGAAATCAGCTCTGGAAACAAATGGAGCGAAGCTTCTGCAG 1172  
Qy 301 ArgTyrAspGlnLeuMetGluAlaLeuGlnLysLysValGluArgIleGluAsnAspPro 320  
Db 1173 CGCTATGACAGGCTCATGAGAGGCTGTGGAGAAAGAGTGGAGCGCATCGAAGACACCCC 1232  
Qy 321 ArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340  
Db 1233 CGGCGCGGGCCCAAGGAGACCAAGTCCGCGAGTCACTACAGAAAGCGTTCCCTGAGATC 1292  
Qy 341 ArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGly 360  
Db 1293 CGCAGAGCGCGGCTGACAGAGGCGATCAGACAGAGGAGGAGGCGGAGCGGAGAGGGG 1352  
Qy 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380  
Db 1353 CTGTCCATGTCCGCGCGCCGACGAGCAGAGGTGTCAAGATCATCATGCTGCTCTCA 1412  
Qy 381 GluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400  
Db 1413 GAGCAGAGAGAACTTGAGAAAGACAGATGCGCAGCTGCGCTGATCCCGCCATGCTGAC 1472  
Qy 401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420  
Db 1473 GAGCTGACACAGAGCCCATCAAGTTCATCAACATGAACGCGCTTATGCGGACCCATG 1532  
Qy 421 LysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPhe 440  
Db 1533 AAGGTGACAAACACCGCCAGTCAATGACATGAGATGAGACAGAGAAAGAGACCTTC 1592  
Qy 441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460  
Db 1593 CGGAGAGATTCATGACATCCCAAGAACTTGTGCTGATCGCATCTCTGAGAGAGG 1652  
Qy 461 LysThrValAlaGluCysValIleuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480  
Db 1653 AAGACAGTGGCTAGAGCGCTCTTATTACTTGACTGAATAAGAAATGAAGAACTATAG 1712  
Qy 481 SerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGln 500  
Db 1713 AGCTGTGAGAGAGAGACTATCGGCCCGGCAAGAGCTGCTGCTGCTAAAGTTCAG 1772  
Qy 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520  
Db 1773 CAAAAAAGATCACCGCGCGAGATGAAGAGCTCCAGAGAGGTTCTGCGCCAGAAAG 1832  
Qy 521 AspGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLys 540  
Db 1833 CCACAGCGCAGAAAGAGCGCTGCTCCAAAGACTCAGAAAGGTCAAAAAGCTCCAGCCC 1892  
Qy 541 AspLysGlu-----AspLeuLeuLysGluLys 549  
Db 1893 AGAAAGCACTGCTCCAAAGGCACTGCGCAAGAAACATAAGTGCATATAAAAAAGTA 1952  
Qy 544 -----AspLeuLeuLysGluLys 549  
Db 1953 ATAAAGTCTTTTGTGACCTGTTAAAAAAA 1985
```

RESULT 5
BC020427 1362 bp mRNA linear HTC 03-JAN-2002
LOCUS Homo sapiens, clone IMAGE:4179307, mRNA.
DEFINITION BC020427
ACCESSION BC020427.1 GI:18043012
VERSION
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (baaes 1 to 1362)
Strausberg, R.
Direct Submission
Submitted (03-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: villalobdbcm.tmc.edu
Villalob, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gunaratne, P., Yoon, V., Kowis, C., Martin, R.,
Lawrence, S., Richards, S., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAK Plate: 12 Row: 1 Column: 23
This clone has the following problem: no polyA-tail.
Location/Qualifiers

FEATURES

source

1. 1362
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4179307"
/tissue_type="Brain, anaplastic oligodendroglioma with
1p/19q loss"
/clone_lib="NCI CGAP_Brn67"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT6"

ORIGIN

Alignment Scores:

Pred. No.: 1,34e-122 Length: 1362
Score: 2264.00 Matches: 431
Percent Similarity: 99.54% Conservative: 0
Best Local Similarity: 99.54% Mismatches: 2
Query Match: 17.13% Indels: 0
DB: 3 Gaps: 0

US-09-522-753-5 (1-2517) x BC020427 (1-1362)

QY 1676 ThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMet 1695
Db 63 ACGGCGCGCTGGAGAACCGGAGACCATCATATGACATACATCCTCGGAGCATG 122	
QY 1696 HisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerPro 1715	
Db 123 CACCAACAACGCGGCCACCGGCATGGCCAGCGAGCTGATGCTGAGGGGCTCTGCC 182	
QY 1716 ArgGlnSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSer 1735	
Db 183 CGGAGTCTCTGGGACTCAACTACGCTGGCGGGTCCCGAGGACATCATGACTGTCC 242	
QY 1736 GlnValProHisLeuProValLeuValProProThrProGlyThrProAlaThrAlaMet 1755	
Db 243 CAAGTCCACACCTGCTGTGCTGTGCTGCCCGACACACGACGCCAGCCGCGCATG 302	
QY 1756 AspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSerSer 1775	
Db 303 GACCGCTTGCTTACCTCCACCGCGGCCACCCCTTTCAGACGCGCCACAGAGCTCC 362	
QY 1776 ProLeuSerProGlyGlyProThrHisLeuThrIleProThrThrThrSerSerSerGln 1795	
Db 363 CCACTCTCCCAAGAGGTCCAAACACTTGAACAAACCAACCAACAGTCTCTGTCGAG 422

QY 1796 ArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGlyLeuSerIleLeu 1815
Db 423 CGGAGCGGAGACCGGGATCGAAGCGGGACCGGGATCGGGAGAGGAGAAAGTCCATCC 482	
QY 1816 ThrSerThrThrThrValGluHisAlaProIleTTPArgProGlyThrGluGlnSerSer 1835	
Db 483 AGTCCACACGACCGGTGAGACCGACCATCTGGAACCTGGTACAGACGAGACGAC 542	
QY 1836 GlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSer 1855	
Db 543 GCGAGAGAGCGAGACGCGGGGGGTGGGGGACAGACGCGCGCCGCTCCCATCC 602	
QY 1856 HisAlaHisGlnHisSerProIleSerProAlaGlnHisAlaLeuGlnGlnArgPro 1875	
Db 603 CATGCCACACACACTCGCCCATCTCCCTCCGAGCCAGAGATGCTCCACAGAGAGCC 662	
QY 1876 SerValLeuHisAsnThrGlyMetLeuGlyIleIleThrAlaValGluProSerLeuPro 1895	
Db 663 AGTGTCTTCACACACGACGATGAAGGATATCATCCGCTGTGAGGCCACGACGCC 722	
QY 1896 ThrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaThrPhePro 1915	
Db 723 ACGTCTCGAGGTCCACCTCCACCTCTCCGCGGCTGCGCCATTCACCT 782	
QY 1916 AlaThrHisCysPProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluPro 1935	
Db 783 GCACCCACTGCGCCACTGGCGGCGGACCCCGATGGGGGTCTACCTCATGAGGCC 842	
QY 1936 ValLeuLeuProGlyGluAlaProArgValAlaArgProGluArgProArgAlaAspThr 1955	
Db 843 GCTTGCTGCGCCAGAGAGGCCCCCGGGTCCCGGCGGACGAGCGGCCGAGCAGAC 902	
QY 1956 GlyHisAlaPheLeuAlaLeuAspProAlaArgSerGlyLeuGluProAlaSerSerPro 1975	
Db 903 GGCATAGCTTCTCTGCGCCAGGCCCGCCGCTCGGGCTGAGGCCGCTCTCC 962	
QY 1976 SerLeuGlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAla 1995	
Db 963 ACGAAGGCTCGAGAGCCCCCGGCTAGTCTCTCTGCGCCACGCGCCACCATGCGC 1022	
QY 1996 ArgThrProAlaAlaAsnLeuAlaProHisHisAlaSerProAspProProAlaProPro 2015	
Db 1023 CGACACCTCGCAAGAACTCGACCTCAACAGCCAGCCCGGACCGCGGCGCACCT 1082	
QY 2016 AlaSerAlaSerAspProHisArgGluThrGlnSerLeuPheSerIleGlnGlu 2035	
Db 1083 GCTTGGCTCGGACCGGACCGGAGAAAGACTCAAAAGTAAACCTTTCCATCCAGAA 1142	
QY 2036 LeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluPro 2055	
Db 1143 CTGGAATCTCGTTCTCTGGGTTTACACGCGACAGCTACAGCCCCGAAAGGGGTGAGGCC 1202	
QY 2056 ValSerProValSerSerProSerLeuThrHisAspLeuGlyLeuProProHisLeuGlu 2075	
Db 1203 GTCAAGCTGTGAGCTCACCACTGACCCAGCAAGAGGGGTCTCCCAAGCACTGGAA 1262	
QY 2076 GlnLeuAspLeuSerHisLeuGluGlyGluLeuArgProLeuGlnProGlyProValLys 2095	
Db 1263 GAGCTGCAAGAAGCACTTGAAGGGGAGCTGCGGCCAAGCAGCAGAGCCCGGTGAAG 1322	
QY 2096 LeuGlyGlyGluAlaHisLeuProHisLeuArgPro 2108	
Db 1323 CTTGGCGGGAGGCGCCCACTCCACACACTGCGGCGC 1361	
RESULT 6
BC026028 1966 bp mRNA linear HTC 22-MAR-2002
LOCUS
DEFINITION Homo sapiens, similar to nuclear receptor co-repressor 1, clone
IMAGE:3836580, mRNA.
ACCESSION BC026028.1 GI:19664000
VERSION
KEYWORDS HTC.


```

SOURCE      Homo sapiens (human)
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1966)
AUTHORS     Strausberg,R.
TITLE       Direct Submission
JOURNAL     Submitted (20-MAR-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT      Contact: MGC help desk
            Email: gcgpb@remail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chin, Chris Fell, Erin Garland, Kan Guin,
            Letticia Hsieh, Martin Krzywinski, Reta Kutische, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Nees, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Diane Smilgus, Michael Smith, Lorraine Spence, Jeff Stoltz,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

FEATURES
source
1..1966
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cisone="IMAGE:3836580"
/cisone_lib="Ovary, adenocarcinoma"
/lab_host="DH10B-R"
/note="Vector: pOTB7"

ALIGNMENT SCORES:
Pred. NO.:          5.95e-86           Length:          1966
Score:              1643.50           Matches:           343
Percent Similarity: 73.35%           Conservative:      56
Best Local Similarity: 63.05%         Mismatched:        98
Query Match:        12.44%           Indels:           47
DB:                 3                Gaps:             10

US-09-5322-753-5 (1-2517) x BC026028 (1-1966)
QY 16 GIUFRoArqTYrPrOPRoHiSeSerIeuSeTyRProVaIGniIleaJaThNiEtHr 35
Db 392 CAAAGTCGTATCTCTCCACATCTGTCACAGTATACATTCCCAACACCSCCASCACAG 451
QY 36 AApVAalgyLeuLauGuIuTYrGLn----hiSiSeRaIgaerTYrAlaSeRhieIeu 53
Db 452 GAGTTCCGACGGCCCGATTAATCGTTCCTCATCTTGGAAGTAGAGTCAGGCCATCACAGCTT 511
QY 54 SeRProgiySeRIlelGIaInProGIaRghRaRgrProSeRleuLauSeRgiuPhgeIN 73
Db 512 TTGCAGCAACGACGACGACAAACAGCTTGAAAGGSGAACCTTCCTTGACAAATTCAC 571
QY 74 ProGIaSnGIuaRgSeRGIuNIleuNihVleuAhProgiuSeRhiiSeRTYleuPro 93
Db 572 CCAGGTTCTGACAGGCGCTCAAGAA-----AGAGAAGACTAGTATGAAACGGTTTCATCCA 625

```

QY	94	GIuLeuGlyLysSerGIuMeGIuPheIleGIuSerLysArgProArgLeuGIuLeuLeu	113
Db	626	GGCCCATCCCAAGTGGAATCATGATTCACCTGGAAATGAAAGCGAACAGCTTGGAACAGGT	685
QY	114	ProAEP-----ProLeuLeuArgProSerPro	122
Db	666	TCTGATTTCTCATTTTCACGGCTCAGTGTCTGGGCTTTTCCCTTTAGTGCACCCGCTGCCA	745
QY	123	LeuLeuAlaThrGIuGlnProAlaGIuSerGIuAEPLeuThrLysAspArgSerLeuThr	142
Db	746	-----GAAAGGCTGAGGGCT---TCTGCAGATGTCAAGAAAGATCCACGATTCGGCA	793
QY	143	GIuLysLeuGIu---ProValSerProProSerProProHisThrAspProGIuLeuGIu	161
Db	794	GGCAAAACATGAAAGCTCCATCTCTCCATATTCGGGGCAACCATGTGAGATATCAAAAT	853
QY	162	LeuValProProArgLeuSerLysGIuLeuLeuIleGlnAsnMetAspArgValAspArg	181
Db	864	GCTTCACCTTCAAAACCTTCAAAGGAAAGATTAAACAGAGTATGGATCTGTAGATCCA	913
QY	182	GIuIleThrMetValGIuGIuGlnGlnIleSerLysLeuLysLysLeuGIuGlnGIuLeuGIu	201
Db	914	GAAATTCGAAAGTAGTAAGACAGACAGATCTTAACTGAAAAAAGAACAAACACGCTGGA	973
QY	202	GIuGIuAlaAlaLysProGIuProGIuLysProValSerProProProIleGIuSer	221
Db	974	GAAAGGACAGCTAAACCTCTGAGCTGAGAAAGCCGCTGCCCTCTCTCTGAGCAG	1033
QY	222	LysHisAspSerLeuValGlnIleIleLysThrAspGIuAsnArgLysLysLeuAlaGIuAla	241
Db	1034	AAACACCCCAAGTATGTCCAAATTTATATGATGGAATTCGAAAAAAGCAAGAAAGCT	1093
QY	242	HisArgIleLeuGIuGIuLysGIuProGlnValGIuLeuProLeuTyTAsnGlnProSer	261
Db	1094	CATAAAATTTTGAAGGCTTGCGCCCAAAAGTTGAACCTGCATATTAACCAAGCATCA	1155
QY	262	AspThrArgGlnTyTHisGIuAsnIleLysIle-----	272
Db	1154	GATACCAAGCTGTACCATGAGAACATCAAGACTGGAAGTACCTGCMAAGCGCATGATGAA	1213
QY	273	AsnGlnAlaMetArgLysLysLeuIleLeuTyTPhelyAspArgAsnHisAlaArgLys	292
Db	1214	AACCAAGGTGAAGAGAAAAAACTCATTTTATTTTAAAGAGAAATCATGCAAGAAA	1273
QY	293	GlnTyTPhelyAsnLysPheCysGIuArgTyTAspGlnLeuMetGlnAlaLeuGIuLysLys	312
Db	1274	CAAAAGGAAACAAAAATCTGCACGGTATATATCATGCTCATGAGAGCATGGAGAAAAAA	1333
QY	313	ValGIuArgIleGIuAsnAsnProArgArgArgAlaLysGIuSerLysValArgGIuTyT	332
Db	1334	GTGCAGCAAAATAGAAAAATATCTCTCGAGAGAAAGCTAAAGAAAGCAAAACAAGGAATTC	1393
QY	333	TyrGIuLysGlnPheProGIuIleArgLysGlnArgGIuLeuGIuArgMetGlnSer	352
Db	1394	TATAAAAAGCGTTTCCAGAAATTCGAAAACAAAGAGAACGCAAGAAAGATTTCAG---	1450
QY	353	ArgValGIuGlnArgLysSerGIuLeuSerMetSerAlaAlaArgSerGlnHisGIuVal	372
Db	1451	CGAGTGTGGCAGAGGGAGAGCTGCTTCACCCACCAATTCCTAGGAGTGAGCATAGATT	1511
QY	373	SerGIuIleIleAspGIuLeuSerGIuGlnGIuAsnLeuGIuLysGlnMetArgGlnLeu	392
Db	1511	TCTGAATTAATTAATGGCTCTCTGAGCAGAGAAATATAGAAACAATGCGCAGCTC	1576
QY	393	AlaValIleProProMetLeuTyTAspAlaAspGlnIleArgIleLysPheIleAsnMet	412
Db	1571	TCTGTGATTCACCTATGATGTTTGAATGACAAACAAAGACAGATCAAGTTCAATTAACATG	1633
QY	413	AsnGIuLeuMetAlaAspProMetLysValTyTArgAspArgGlnValMetAsnMetCrr	432
Db	1631	AATGGGCTTAATGAGAGACCCCTATATAAGTATATAAGTATAGCAGGTTTATATAAGTTTGG	1696
QY	433	SerGIuGlnGIuLysGIuThrPheArgGIuLysPheMetGlnHisProLysAsnPheGIu	452

[illegible]

RESULT 7			
LOCUS	BC035748		
DEFINITION	BC035748	1850 bp	mRNA
	Homo sapiens nuclear receptor co-repressor 1, mRNA (CDNA clone IMAGE:5749816), with apparent retained intron.		linear
			HTC 19-NOV-2003

SOURCE	Homo sapiens (human)
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eultheria; Primates; Carnivora; Homnidae; Homo.
1 (bases 1 to 1850)
Srausberg, R. L., Feingold, E. A., Grouse, J. H., Derge, J. G.,

TITLE

human and mouse cDNA sequences

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 1850)
Strausberg, R.
Direct Submission
Submitted (31-JUL-2002) National Institutes of Health, Mammalian

REMARK	COMMENT
NIH-MGC Project URL: http://mgc.nci.nih.gov	
Contact: MGC help desk	

Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc.mc@nih.gov
Akhter, N., Ayte, K., Beckstrom-Sternberg, S. M., Benjamin, B.,
Blakesley, R. W., Boultard, G. G., Breen, K., Brinkley, C., Brooks, S.,
Dierlich, N. L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S. D., Karlins, E., Khong, P., Laric, P., Legaspi, R.,
Maduro, O. L., Mastillo, C., Maskeri, B., Mastriian, S. D., McLooney, J. C.,
McDowell, J., Pearson, R., Stentrop, S., Thomas, P. J., Touchman, J. W.,
Young, C., Vogt, J. L., Walker, M. A., Wetherby, K. D., Wiggins, L.,
Young, A., Zhang, L., H. and Green, E. D.

clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at: <http://image.lnl.gov>
Series: IRAK Plate: 79 Row: m Column: 8
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA bc1: 22538460
This clone has the following problem: retained intron.

FEATURES
Source

```

"organism"="Homo sapiens"
"mol_type"="RNA"
"db_xref"="taxon:9606"
"clone"="IMAGE:5749816"
"tissue_type"="Pancreas, Spleen, adult pooled"
"clone_1fb"="NIR MGC_120"
"lab_host"="DH10B"
"note"="Vector: pCMV-SPORT6"

```

ORIGIN

Alignment Scores:	
Pred. No.:	1,97e-85
Score:	1634.00
Percent Similarity:	74.35%
Best Local Similarity:	63.84%
Query Match:	12.36%
DB:	3
US-09-522-753-5 (1-2517) x BC035748 (1-1850)	
	Length: 1850
	Matches: 346
	Conservative: 57
	Mismatches: 96
	Indels: 43
	Gaps: 12

Qy	16	GIIPProAaGTYrProProHiISerLeuSerYrProValGlnIleAlaGTHrHISThr	35
Db	272	CAAAgTCGTAATCCCTCCCTCACTCTGTCCAGATACATTTCCCAACACCCGCCACAGAC	331
Qy	36	AspValGlyLeuLeuGlyTyrln-----HisHisSerArgAspTyrAlaSerHisIleu	53
Db	332	GAGTTCGACGTCCTCGATTATTCCTTCCTCATCTTAAAGTGAATCAGGCATCAACGCTT	391
Qy	54	SerProGlySerIleIleGlnProGlnArgArgProSerLeuSerGlnupRegln	73
Db	392	TTGCAGCAACAGCAGCAGCAACAGCTTGAAAGGAGACCTTCCTTGCTTTCAGAAITTCAC	451
Qy	74	ProGlyAsnGluArgSerGlnGluLeu---His-LeuArgProGluSerHis-----Se	90
Db	452	CCAGGTTTCGACAGGTTTATTCTTTCTTGATGATTAGGCTTCAGAAAGGAAACTAG	511
Qy	90	rTyrlenPro--GluLeuGlyIlySerGluMetGluPhe-----IleIleuSerIysr	107
Db	512	TTATGAACCGTTTCATCCAGGCCCATCCCGAGTATCATGATTCAGTGAATCGAAGCG	571
Qy	107	gProArGluGlnLeuLeuProAsp-----	116
Db	572	ACCACGCTCGAACAACAGGTTTCGATTCATTTTCACGGTCACTGCTCGGTTTTGCC	631
Qy	116	oLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuTh	136
Db	632	TTTAGTGACCCGCGTCCA-----GAAGGCGTGAAGGCT---TCTCAGATGCTAA	679
Qy	136	rIysAspArgSerLeuThnGlyIlyLeuGlu---ProValSerProSerProProHi	155
Db	680	GAAGATTCACGATTTGAGAGCAACATGAAGCTTCATCTCTTCATTTTCGGGGCAAC	739

```

Qy 155 5THA8PProGluLeuGluValProProA8gLeuSerIySgIuLeuIleGlnAs 175
Db 740 ATGTGAGATGCAAAATGCTTCACCTTCAAACTCTCAAGGAAGATTAATACAG 799
Qy 175 nMeLA8PArGValA8PArGValIleThMeCValGluGlnGlnIleSerIySgIy 195
Db 800 TATGATCGTGTAGATGAGAAATTCAGAAAGTAAGAACAGATCTTAACTGAAAA 859
Qy 195 8LySGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGIn 215
Db 860 GAAACAACACACCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 919
Qy 215 rProProProIleGluSerIySHa8gSerLeuValGlnIleIleTy8PArGInA8 235
Db 920 CCTCCCTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 979
Qy 235 gLy8Ly8ValAGValA8IleA8IleA8IleA8IleA8IleA8IleA8IleA8 255
Db 980 GAAAAAGCAGAAAGAGCTCATAAATTTTGAAGGCTTGGCCCAAAAGTTGAAC 1039
Qy 255 0LeuTy8PArGInPro8e8PArGInTy8Iy8Iy8Iy8Iy8Iy8Iy8Iy8Iy 275
Db 1040 ACTGTATACCAACCATCAGATACCAAGGTGTACATGAGAGAGAGAGAGAGAG 1099
Qy 275 aMeLA8gLy8LeuIleLeuTy8PheTy8ArGArGAsnH8A8IleA8gLy8In 295
Db 1100 GATGAGAGAAAACTCATTTTATTTTAAAGAGAAATCATGCAAGAGAAACAA 1159
Qy 295 8GInLy8PheCy8GInArGTy8PArGInLeuMeGInA8IleuGInLy8Ly8Val 315
Db 1160 ACAAAAATTCGACAGGTATATCATCGCATGAGAGAGAGAGAGAGAGAGAGAG 1219
Qy 315 gIleGInA8PArGArGArGArGArGArGArGArGArGArGArGArGArGAr 335
Db 1220 AATGAAATTAATCTCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1279
Qy 335 8GInPheProGInIleArGLy8GInArGInGInGInGInGInGInGInGInG 355
Db 1280 GCAATTCAGAAATTCAGAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1336
Qy 355 yGInArGLy8SerIyLeuSerMetSerIleA8IleA8gSerIyH8IleGIn 375
Db 1337 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1396
Qy 375 eIleA8P8Ly8LeuSerGInGInGInGInGInGInGInGInGInGInGIn 395
Db 1397 TATGATGGGCTCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1456
Qy 395 eProProMeLeuTy8PArA8IleA8P8GInArGInIleTy8PheIleA8Me 415
Db 1457 TCCACCTATGATGTTGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1516
Qy 415 uMeLA8PArGProMeLeuValIy8Ly8A8PArGInValMeCAsnMeCTrP8 435
Db 1517 TATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1576
Qy 435 nGInLy8GInTy8PheArGInLy8PheMeGInH8IleProLy8A8PheGIn 455
Db 1577 TGAAGAGAGATCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1636
Qy 455 a8ePheLeuGInArGLeuTy8PheValAGInCy8ValLeuTy8Ty8Iy8Th 475
Db 1637 ATCATATCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1696
Qy 475 8A8GInA8Iy8Ty8Ly8SerLeuValArGArG8eTy8Ty8ArGArGInLy8 494
Db 1697 AATGAGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1756
Qy 494 nGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGIn 514
Db 1757 GCAA-----ATTGCTCG 1768
Qy 514 g8e8e8e8e8e8e8e8e8e8e8e8e8e8e8e8e8e8e8e8e8e8e8e8e8 534

```

```

Db 1769 ACCCTCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1828
Qy 534 uLy8 535
Db 1829 AAAA 1832

RESULT 8
BC029627
LOCUS
DEFINITION Mus musculus, similar to nuclear receptor co-repressor 1, clone
ACCESSION BC029627
VERSION BC029627.1
KEYWORDS GI:20988194
SOURCE
ORGANISM Mus musculus (house mouse)
REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (06-MAY-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
COMMENT NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabos-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Guanarane, P.H., Garcia, A.M., Lu, X., Huylk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNLN at: http://image.llnl.gov
Series: IRAK Plate: 31 Row: K Column: 14
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, similarity but not identity to protein
This clone has the following problem: frame shifted.
Location/Qualifiers
1. 1835
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:4501757"
/issue_type="Eye, retina, mouse strain C57BL/6"
/clone_lib="NIH_MGC_94"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

ALIGNMENT SCORES:
Pred. No.: 5,37e-85 Length: 1835
Score: 1626.50 Matches: 342
Percent Similarity: 73.43% Conservative: 56
Best Local Similarity: 63.10% Mismatches: 93
Query Match: 12.31% Indels: 51
DB: 3 Gaps: 12

US-09-522-753-5 (1-2517) x BC029627 (1-1835)
Qy 16 GluProArGTy8ProProH8e8e8e8e8e8e8e8e8e8e8e8e8e8e8e8 35
Db 282 CAAGTCGTTATCTTCACATTCCTGCGAGTATACCTTCCAGTACCGAGACAG 341

```

```

QY 36 AspValGlyLeuLeuGluTyrGln-----HisHisSerArgAspTyrAlaSerHisLeu 53
   ::: :::::
Db 342 GAAATTGGACGTTCCCTACCTACCGCTCTTCATATGAAAGTTGACGAGCATCAGCTT 401
QY 54 SerProGlySerIleIleGlnProGln---ArgArgArgProSerLeuSerGluPhe 72
   ||| |||
Db 402 TTACAGCAGCAGCAGCAGCAGCAGCTTCGAAAGACGACCTTCCTTGCTTCAGAAATT 461
QY 73 GlnProGlyAsnGlnArgSerGlnGluLeuHisIleLeuArgProGluSer---HisSerTyr 91
   ||||| :::::
Db 462 CACCCGGGTTCTGACAGGCTCAGAAAGAGAACTGATATGACAGTTTCACTCA--- 518
QY 92 LeuProGluLeuGluGlyLeuSerGluMetGluPhe-----IleGluSerIleArgProArg 109
   ||| ||| :::::
Db 519 -----GGACCCCTCACCGCGTGATCATGATTCCTGTGAGTCCAAAGCCCAACCC 566
QY 110 LeuGluLeuLeuProAsp-----ProLeuLeu 118
   ||||| :::::
Db 567 CTGAGACCCAGTTTCTGATGCCCATTTCCAGCGTGTAGTGCCTGGCTTTTACCTTTAGT 626
QY 119 ArgProSerProLeuLeuAlaThrGlyGlnProAlaGly-----SerGluAspLeu 135
   ||| |||
Db 627 CAC-----TCGCTGCCAAGAGCTTGAGGTCTGTCAGATGCT 665
QY 136 ThrLeuAspArgSerLeuThrGlyLeuLeuGlu---ProValSerProProSerProPro 154
   ||||| :::::
Db 666 AAGAAGATTCAGCATTTGGAGACAAACATGAAAGCTCCATCCTCTCTTGCGTGAGCAA 725
QY 155 HisThrAspProGluLeuGluLeuValProProArgLeuSerIleGluLeuIleGln 174
   ||| ||| :::::
Db 726 CCATGTGAGATGACCAAAATGCTTCACCTTCACAAAGAGAGGATTAATACAG 785
QY 175 AsnMetAspArgValAlaAspArgGluIleThrMetValGluGlnGlnIleSerIleLeu 194
   ||||| :::::
Db 786 AGTATGACCGGGTTAGACGAGGATTGCAAAAGTACAGCAGCATTCCTTAATTGAAA 845
QY 195 LysLeuGlnGlnGlnLeuGlnGluGluAlaAlaLysProProGluProGluLysProVal 214
   ||||| :::::
Db 846 AAGAACAACAACAAGCTAGAGAAAGAGAGAGCTAAGCCCTCAGACAGAGAGAGCTGTG 905
QY 215 SerProProProIleGluSerIleHisIleArgSerLeuValGlnIleIleTyrAspGluAsn 234
   ||||| :::::
Db 906 TCCCTCTCTCTCTGTGAGCAAAACACCGTAGTATTGTCCAAATTTATATGATGAAAT 965
QY 235 ArgGlyLysAlaGluAlaHisArgIleLeuGluGlyLeuGlnArgProGlnValGluLeu 254
   ||||| :::::
Db 966 CGAATAAAGCAGAGAGAGCTCATAAATATTTGAAGTCTTGCCCAAAAGTTGAACGTG 1025
QY 255 ProLeuTyrAsnGlnProSerAspThrArgGlnTyrHisIleGluAsnIleLysIleAsnGln 274
   ||||| :::::
Db 1026 CCACCTATATACAGCAGCATAGATCCAAAGGTCTACATAGAACATCAAGACAAACCGAG 1085
QY 275 AlaMetArgLysIleLysLeuIleLeuTyrPheLysArgArgAsnHisAlaIleArgLysGlnTyr 294
   ||||| :::::
Db 1086 GTGATATAGGAAAAAATCATTTATTTTAAAAAGAAATATCATCAAGAAAAACAAAG 1145
QY 295 LysGlnLysPheGlyGlnArgTyrAspGlnLeuMetGluAlaLeuGlnLysLysValGln 314
   ||||| :::::
Db 1146 GAACAAAAAATCTGCGAGCGCTATGATCAGCTCATGAGACCGGAGGAAAAAAGTGAGAC 1205
QY 315 ArgIleGluAsnAsnProArgArgArgAlaLysGluSerLysValAlaArgLysTyrTyrGln 334
   ||||| :::::
Db 1206 AAGATAGAAAAATATCCCGGAGAGAAACGAAAAAGAAACAAACAGGAAATCTATGAG 1265
QY 335 LysGlnPheProGluIleArgLysGlnArgGluLeuGlnGlnLysArgMetGlnSerArgVal 354
   ||||| :::::
Db 1266 AAGCACTTTCAGAAAAATTCAGAAAAACAAAGAGAAACAGAGAAAAAGATTTCAG--CGAGTT 1322
QY 355 GlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGlnHisGlnValSerGln 374
   ||||| :::::
Db 1323 GGTCAAGGGGAGACTGCTTTTCAGCCACCATTTGAGAGTAGAGCATGAGATTTCGAA 1382

```

```

QY 375 IleIleAspGlyLeuSerGlnGlnGlnLysLeuLeuGlnLysGlnMetArgGlnLeuAlaVal 394
   ||||| :::::
Db 1383 ATTATGATGCTCTTTCTGAAACAGAGAAATATGAGAAAGCAAAATCGCTACGCTTCTGTG 1442
QY 395 IleProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetArgGly 414
   ||||| :::::
Db 1443 ATTCAACCTATATGATGTTGATCAAGAACAAAGAGGTCAAATTCATCAATATGAAATGG 1502
QY 415 LeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTyrSerGln 434
   ||||| :::::
Db 1503 CTGATGAGAGATCCCATATGAGCTTTATTAAGACAGACAGATTATGATGTTTGGACTGAC 1562
QY 435 GlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIle 454
   ||||| :::::
Db 1563 CATGAAAAAGAGATCTTTTAAGACAAAGTTTATCCAGCATCCAAAAAATCTTGACTAAAT 1622
QY 455 AlaSerPheLeuGlnAlaArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLys 474
   ||||| :::::
Db 1623 GCATCCTATTTGAAAAGAGAGGTCTCTGATGTGTTTATATTAATTAATTAACCAAG 1682
QY 475 LysAsnGluAsnTyrLysSerLeuValArgArgSerTyr---ArgArgArgGlyLysSer 493
   ||||| :::::
Db 1683 AAAAATGAAATTTATTAAGGCCCTCGTAGAAGAGATTTAGAAAAACGAGAGCAAGAAAT 1742
QY 494 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 513
   ||||| :::::
Db 1743 CAGCAG-----ATTGCC 1754
QY 514 ArgSerSerGlnGlnGluLysAspGluLysGluLysGluLysGluAlaGluLysGluGln 533
   ||||| :::::
Db 1755 CGTCCCTCAAGAAAGAAAGAAAGTAAAGAAAGAAAGAGATTAAGCAGAAAAACAGAA 1814
QY 534 GlnLys 535
   ||||| :::::
Db 1815 AAAAAA 1820

RESULT 9
BC062162 1819 bp mRNA linear HTC 14-NOV-2003
LOCUS
DEFINITION
Mus musculus cDNA clone IMAGE:30240594, containing frame-shift
errors
ACCESSION
BC062162
VERSION
BC062162.1 GI:38328289
KEYWORDS
HTC.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1819)
REFERENCE
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shennan,C.M., Schuler,G.D.,
Altshul,S.F., Zeeberg,B., Buetow,K.H., Scheffer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Haieh,F.,
Diachenko,L., Marusina,K., Farmer,A.A., Rudin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,F.L.,
Scheetz,T.E., Brownstein,M.J., Ueda,T.B., Toshituki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Millahy,S.J., Bosak,S.A., McMan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Huliyil,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahy,J., Helton,E., Kettelman,M., Madan,A.C., Rodrigues,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shchepochko,Y.,
Bouffard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmitz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalins,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J., and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
JOURNAL
MEDLINE
PUBMED
22388257
12477932
REFERENCE
2 (bases 1 to 1819)

```

AUTHORS Straubeberg, R.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Dr. David Rowe and Dr. Mina
 CNA Library Preparation: Invitrogen Corp
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
 BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Kirzysinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McAvery, Steven Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalins, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natesja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAC Plate: 131 Row: c Column: 2
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, similarity but not identity to protein
 This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers
 1. 1819
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:30240594"
 /tissue_type="Jaw bone, mouse, day 10.5 to 11.5 (CD-1), developing maxilla and mandibula tissue containing undifferentiated progenitor cells for muscle, dermis, epidermis, skin, membranous bone, cartilage and teeth"
 /clone_lib="NIH_MGC_164"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6.1"

ORIGIN

Alignment Scores:

Pred. No.:	Length:	1819
Score: 1624.50	Matches: 341	
Percent Similarity: 73.94%	Conservative: 59	
Best Local Similarity: 63.03%	Mismatches: 91	
Query Match: 12.29%	Indels: 50	
DB: 3	Gaps: 11	

US-09-522-753-5 (1-2517) x BC062162 (1-1819)

QY 16 GlnProArgGlyProProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThr 35
 Db 282 CAAAGTCGATCTTCATCTGTCACAGTATACCTTCCACGATACCGACACGACG 341
 QY 36 AspValGlyLeuLeuGlnTyrGln-----HisHisSerArgAspTyrAlaSerHisLeu 53
 Db 342 GAATTTGCAAGTTCCTGACCTACCGCTCTTCATATTGAAGTTAGCCAGCATACAGCTT 401
 QY 54 SerProGlySerIleIleGlnProGln--ArgArgArgProSerLeuLeuSerGluPhe 72
 Db 402 TTACACAGACGACGACGACGACGACGCTTCGAAAGCGACCTTCCTTGCTTACGATTT 461
 QY 73 GlnProGlyAsnGlnLysSerGlnGlnLeu-HisLeuArgProGluSerHisSerTyrIle 92
 Db 462 CACCCGGGTTCTACACGGCTCAAGAAAGAGAACTGAGTAAACAGTTTACTCA----- 516
 QY 92 uProGluLeuGlyLysSerGluMetGluPhe-----IleGluSerLysArgProArgLe 110

Db 517 -----GACCCCTCACCGGATCATGATTCCTTGAGTCAAGGCCACGCT 566
 QY 110 uGluLeuLeuProAsp-----ProLeuLeu 119
 Db 567 GAGACGATTTCTGATGCGCATTTCCAGCGTGTAGTCTGCGGTTTACCTTATGTTCA 626
 QY 119 gProSerProLeuLeuAlaThrGlyGlnProAlaGly-----SerGluAspLeuThr 136
 Db 627 C-----TCGCTGCCAAGAGCTTGAAGCTTGAAGCTCTGCATGATGCTTA 665
 QY 136 rLysAspArgSerLeuThrGlyLysLeuGlu---ProValSerProSerProProHis 155
 Db 666 GAAGSATTCAGCATTTGGAAGCAAAACATGAAGCTCCATCTCTCTTGGCTGGCAACC 725
 QY 155 sThrAspProGluLeuGlnLeuValProProAlaGlyLeuSerGlyGlnLeuIleGlnAs 175
 Db 726 ATGTGAGATGACCAAAATGCTTCACCTTCAAGCTTCAAGGAGATTAATATACAG 785
 QY 175 nMetAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLys 195
 Db 786 TATGACCGGTTAGACCGAGAGATTGCAAAAGTACAGCAGATCTTAAATTGAAAA 845
 QY 195 sLysGlnGlnLeuGlnGlnGlnIleAlaLysProProGluProGluLysProValSe 215
 Db 846 GAAACAAACACAGCTAG 905
 QY 215 rProProProLysGluSerLysHisArgSerLeuValGlnIleIleLysArgGluAsn 235
 Db 906 CCCTCTCTCTGAGCAAAACCAAGGATATGTCAAATTAATTAATATGAGAAATCG 965
 QY 235 gLysLysAlaGlnAlaAlaHisArgIleLeuGlnGlyLysGlyProGlnValGluLeuPr 255
 Db 966 GAAAAAGCAAGAAAGCTATTAATATTGAAAGCTTTGGCCCAAAATTTAACTGCC 1025
 QY 255 oLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGln 275
 Db 1026 ACTGTAATACACAGCATCAGATGACCAAGGTGACATGAGACATCAACAGCAACGAGT 1085
 QY 275 sMetAspGlyLysLeuLeuLeuLeuTyrPheLysValGlnGlnAlaAlaArgLysGlnTrp 295
 Db 1086 GATGAGAAAAAACTCATTTTATTTTAAAGAAATCAATGACCAAGAAACAAAGGGA 1145
 QY 295 sGlnLysPheCysGlnArgTyrAspGlnLeuMetGlnAlaLeuGlnLysValGluAr 315
 Db 1146 ACAAAAATCTGCGACGCTATGATCATGACCTCATGAAAGCTGGAGAAAAATGACAG 1205
 QY 315 gIleGluAsnAsnProArgArgArgAlaLysGlnSerLysValArgGlnTyrTyrGlnLys 335
 Db 1206 AATGAAATTAATCCCGGAGGAAACCAAAAGAAACCAAAAGGAAATACATGAGAA 1265
 QY 335 sGlnPheProGluIleArgLysGlnArgGlnLeuGlnGlnLysArgValGln 355
 Db 1266 GCAGTTTCCAGAAATTCGAAACCAAGAGAACGACAAAGAAAGATTTTCAG---CGAGTTGG 1322
 QY 355 yGlnArgGlySerGlyLeuSerMetSerAlaAlaAspSerGlnHisGluValSerGlu 375
 Db 1323 TCAGAGGAGAGCTGCTTTTACGCCACCATGCTTASAGATGACATGAGATTTCTGAAT 1382
 QY 375 eIleAspGlyLeuSerGluGlnGlnLeuAsnGlnLysGlnMetArgGlnLeuAlaVal 395
 Db 1383 TATTGATGCTTTTTCGACAGAGAAATATGAGAAATGCAATGCTGCTCTGAT 1442
 QY 395 eProProMetLeuTyrAspAlaAspGlnGlnIleValIleLysPheIleAsnMetAsnGly 415
 Db 1443 TCACCTATGATGTTTGAAGCAAGAAAGAGGCAAAATTCATCAATATGATGATGGCT 1502
 QY 415 nMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGlu 435
 Db 1503 GATGAGAGATCCAAATAGGTTTATTAAGACACAGCTTATGAACTTTGAGCTAGCA 1562
 QY 435 nGluLysGlnThrPheArgGluLysPheMetGlnHisProLysValAsnPheGlyLeuIle 455

QY 2486 agly---SerGlyProLeuAla---GlyProHis-HisAlaATP 2498
DB 1039 CGGGGAAGCGGGCCCTTCGCGTGGCCCCCAGCCAGCGCTGG 1082

RESULT 11
LOCUS BC026623
DEFINITION Mus musculus, similar to nuclear receptor co-repressor 1, clone
IMAGE:4981672, mRNA.
ACCESSION BC026623
VERSION BC026623.1 GI:20073007
KEYWORDS HTC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1996)
REFERENCE 1
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (02-APR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.sbgc.stanford.edu>
Contact: (Dickson, Mark) medpaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRM Plate: 59 Row: m Column: 5
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis
This clone has the following problem: frame shifted.
Location/Qualifiers
1..1996
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4981672"
/rname="Colom, normal. 5 month old male mouse."
/clone_id="NCI CGAP_CO24"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

ORIGIN
Alignment Scores:
Pred. No.: 1.45e-83 Length: 1996
Score: 1603.00 Matches: 337
Percent Similarity: 73.40% Conservative: 52
Best Local Similarity: 63.58% Mismatches: 91
Query Match: 12.13% Indels: 50
DB: 3 Gaps: 12

US-09-522-753-5 (1-2517) x BC026623 (1-1996)
QY 16 GIUProArgTyProHisSerLeuSerTyProValGlnIleAlaArgThrHisThr 35
DB 490 CAAAGTCGTTATCTTCACATCTGTGCGATATACCTTCCAGTACCGACACGACG 549
QY 36 AepValGlyLeuLeuGlnTyArgin-----HisHisSerArgAspTyAlaSerHisLeu 53
DB 550 GAATTTCAGTTCCTGATACCGCTCTTCATATTGAAGTTAGCCAGGATCATCAGCTT 609

QY 54 SerProGlySerIleIleGlnProGln---ArgArgArgProSerLeuLeuSerGluPhe 72
DB 610 TTACACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 669
QY 73 GlnProGlyValGlnGlySerGlnGlnLeuHisLeuArgProGluSer---HisSerTy 91
DB 670 CACCCGGGTTCACACGACGCTCAAGAAAGAGAACTGATATGAAACAGTTCACTCA--- 726
QY 92 LeuProGluLeuGlnGlySerGlnMetGluPhe-----IleGluSerTyValArgProArg 109
DB 727 -----GACCCCTCACCGGTGATCATGATTCCTTGAGTCCAGCCGCCACGC 774
QY 110 LeuGluLeuLeuProAsp-----ProLeuLeu 118
DB 775 CTGAGCCAGTTCCTGATGCCCTTCCAGCGGTGATGCTCCGCTTACCTTTAGTT 834
QY 119 ArgProSerProLeuLeuAlaThrGlyGlnProAlaGly-----SerGluAspLeu 135
DB 835 CAC-----TCGCTGCCAAGAGGCTTGAGTGTGTCGACGATGCT 873
QY 136 ThrTyAspArgSerIleThrGlyLeuGln---ProValSerProProSerProPro 154
DB 874 AAGAGGATTCACGATTTGGAGACCAAGTGAAGCTCCATCTCTCTTGGCTGGGCAA 933
QY 155 HisThrAspProGluLeuGlnValProProArgLeuSerTyGlnGlnIleGln 174
DB 934 CCATGTGAGATGACCAAAATGCTTCACTTCAAAAGCTTCAAGAGGATTAATACAG 993
QY 175 AsnMetAspArgValAlaAspArgGluIleThrMetValGlnGlnIleSerTyLeuTy 194
DB 994 AGTATGACCGGGTATACCGAGATTTGCAAAAGTAGAGACAGATCTCTTAATTTGAAA 1053
QY 195 TyLeuGlnGlnGlnLeuGlnGlnGlnAlaAlaTyProProGluProGluTyProVal 214
DB 1054 AAGAAACAACAACAGCTAGAAAGAAAGCAAGTCAAGCCCTTCAAGAGAGCCCTGTG 1113
QY 215 SerProProProIleGlnSerTyHisArgSerLeuValGlnIleTyTyAspGluAsn 234
DB 1114 TCCCTCTCTCTGTGAAACAAACACCGTATGATTTCCAAATTTTATGTAGTGAAAT 1173
QY 235 ArgTyValAlaGlnAlaAlaHisArgIleLeuGlnGlyProGlnValGlnLeu 254
DB 1174 CGGAAAAAGACGAAGAGCTGTAATTAATTTGAAGGTCTTGCCCAAGTTGAACGTG 1233
QY 255 ProLeuTyArgGlnProSerAspThrArgGlnTyHisGluAsnIleTyValGln 274
DB 1234 CCACTCTATACCAAGCATCAGATACCAAGGTATCATGAGAACATCAAGACAAACAG 1293
QY 275 AlaMetArgTyLeuTyLeuTyPheTyArgTyAspHisAlaIleArgTyValGln 294
DB 1294 GTGATGAGAAAAAATCTCATTTTATTTTAAAGAAATCATATCAAGAAACAAAGG 1353
QY 295 TySGlnTyPheCysGlnArgTyArgGlnLeuMetGluAlaLeuGlnTyValGln 314
DB 1354 GAACAAAAAATCGCCAGCGCTATGATGATCATGATGAAAGCGTGGAGAAAAAGTGAAC 1413
QY 315 ArgIleGlnAsnAsnProArgArgAlaAlaTySerTyValArgGlnTyTyGln 334
DB 1414 AGAATGAAAAATATCTCCCGGAGAAAGCAAAAGAAACAAACAAAGGAAATCTATGAG 1473
QY 335 TySGlnPheProGlnIleArgTySGlnArgGlnLeuGlnGlnTyArgMetGlnSerArgVal 354
DB 1474 AAGCAGTTCCAGAAATTCGAAACAAAGAAACAGCAAGAAAGATTTGAG---CAGATT 1530
QY 355 GlyGlnArgTySerGlyLeuSerMetSerAlaAlaArgSerGlnIleGlnValSerGln 374
DB 1531 GGTCAAGAGGAGAGTGTCTTTCAGCCACCATTTGCTGAGAGATGATGATTTCTGAA 1590
QY 375 IleIleAspGlyLeuSerGlnGlnIleGlnLeuGlnIleMetArgGlnIleAlaVal 394
DB 1591 ATTATTGATGATCTTCTTGAAACAGAGAAATTAATGAGAACAAATGCTGAGCTTCTGTG 1650

QY 395 11eProPromeLeuTyRaph1aAepGIngaRgl1eLysPhe11eAsnMetCangly 414
Db 1651 ATTCACCTATGATGTTTATGATGACAAAGAGGCGCAATTATCATATATGAAATGCG 1710
QY 415 LeuMetAlaAspPromeTylsVal1TyRlyAspAArgInValMetAsnMetTpsRglu 434
Db 1711 CTGATGAGAGATCCATGAAAGGTTTAAAGACAGACAGTTTATGATGATGTTGGACTGAC 1770
QY 435 GIngluTyG1uTherPheArgGluTyGlyPheMetGln1eProLysAsnMetGlyLeu1le 454
Db 1771 CATGAAAGAGATCTTTAAGACAAAGTTTATTCAGCATCCAAAACTTGGACATAAT 1830
QY 455 AlaSerPheLeuGluArgTyRThVal1aGluCysValleuTyTyTyRleuThly 474
Db 1831 GCATCTTATTTGAAAGAAAGAGTGTTCTCGATGTTGTTTATATTAATTAATTAACCAAG 1890
QY 475 LysAsnGluAsnTyRlySerLeuValArgArgSerTyR---ArgTyRArgGlyLysSer 493
Db 1891 AAAAATGAAATTTAAAGCCCTCTGAGAGAGAAATTAAGAAACGCAAGGAGCAAAAT 1950
QY 494 GIngluGIngluGIngluGIngluGIngluGIngluGIngluGIngluGIngluGInglu 513
Db 1951 CAGCAGCAG-----ATTGCC 1965
QY 514 ArgSerSerGIngluGluLysAspGluLys 523
Db 1966 CGTCCTCACAAAGAAAAA 1995
RESULT 12
BM555371 965 bp mRNA linear EST 20-FEB-2002
LOCUS BM555371
DEFINITION AGENCOURT 6577786 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5470428
ACCESSION BM555371
VERSION 5', mRNA sequence.
KEYWORDS BM555371.1 GI:18795772
SOURCE EST.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 965)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: DCTD/DRP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.jnl.gov
Plate: L1CM1977 row: d column: 13
High quality sequence stop: 670.
Location/Qualifiers
1. 965
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5470428"
/issue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_41"
/note="Organ: skin; Vector: pOTB1; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGGACGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC library."

Alignment Scores:
Pred. No.: 1,956-81 Length: 965
Score: 1559.00 Matches: 304
Percent Similarity: 95.00% Conservative: 0
Best Local Similarity: 95.00% Mismatches: 11
Query Match: 11,80% Indels: 5
DB: 4 Gaps: 3
US-09-522-753-5 (1-2517) x BM555371 (1-965)
QY 2182 AspleuTyRleuProProAspHisGlyAlaProAlaArgGlySerProHisSerGlu 2201
Db 3 GACCTCTACTCTCCGCCCGGACCATGATGCCCCGGCGGCTCCGCCACAGCGAA 62
QY 2202 GlyGlyLysArgSerProGluProAsnTyRSerValLeuGlyGlyGlyLysArgGly 2221
Db 63 GGGGCGAAGAGTCTCCAGAGCCAAACAGAGTGCGTCTTGCGTGCTGAGGAGCGGT 122
QY 2222 11eGluProValSerProProGluGlyMetThrgluProGlyHisSerArgSerAlaVal 2241
Db 123 ATTGAACTGTGTCCCAACCGAGGGCATGACGAGCCAGGGCAGCTCCCGAGTGCTGTG 182
QY 2242 TyRProLeuLeuTyRArgAspGlyGluGlnThrgluProSerArgMetGlySerLysSer 2261
Db 183 TACCGCTGCTGTACCGGATGGGAGAAACAGACGAGCCAGAGATGGGCTTCAAGTCT 242
QY 2262 ProGlyAsnThSerGlnProProAlaPhePheSerTyRLeuThrgluSerAsnSerAla 2281
Db 243 CAGGCGAACACACAGCAGCGCGCAGCTTCTTACAGAGCTACACGAGCAACTCCGCC 302
QY 2282 MetValLysSerLysGlyGlnGluLeuAsnTyRLeuAsnThHisAsnArgGlu 2301
Db 303 ATGTCTCAAGTCCAAAGAGCAAGATCAACAAAGAGCTGMAACCCACCAACCGAATGAG 362
QY 2302 ProGluTyRAsnHisSerGlnProGlyThrgluPheAsnMetProAlaLeuThrgly 2321
Db 363 CCGTAATCAAAATACGACGAGCTGGAGGAGATCTTAATATGCGCCATCACCGGA 422
QY 2322 ThrGlyLeuMetThTyRArgSerGlnAlaValGlnGluHisAlaSerThrasnMetGly 2341
Db 423 ACAGGCTTATATGACCTATGAAAGCCAGCGGCGAGAACATGCCAGCAACATGGGG 482
QY 2342 LeuGluAlaIle1eArgLysAlaLeuMetGlyLysTyRAspGlnThrgluLysPro 2361
Db 483 CTGAGGCGCATATTAAGAAAGCAGCTCATGGTAAATATGACACAGTGGAGAGTCCCG 542
QY 2362 ProluSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaIleMet 2381
Db 543 CCGCTCAGGCGCAATGCTTTAACTCTGAATGCCAGTCCAGCTCCGCTGCTATG 602
QY 2382 Prol1eThralAlaAspGlyArgSerAspHisThrLeuThSerProGlyGlyGly 2401
Db 603 CCGATTAACCGCTGCTGACCGAGAGTACACACACTCACTCCGCAAGTGGCGGG 662
QY 2402 LysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeu 2421
Db 663 AAGGCCAAGTCTCTGGACAGCCAGCAACCAAAAGCCAAATGCCCGGCCCGCGCTG 722
QY 2422 AlaSerGlyAspArgProProSerValSerSerValHisSerGlnGlyAspCysAsnArg 2441
Db 723 GCATCTGGGAGCCGCGACCTCTGTCTCTCAAGTCACTCGAGAGAGACTGCAACCG 782
QY 2442 ArgThrProLeuThrasnArgValTrpGluAspArgProSerSerAlaGlySerThrPro 2461
Db 783 CCGAGCCCGCTACCAACCGCGTGGAGAGACAGGCCCTCGTCCGACAGGTCCAGGCCA 842
QY 2462 PheProTyRAsnProLeuIleMet-ArgLeuGlnAlaGly---ValMetAlaSerProP 2480
Db 843 TTCCCTCAAAACCCCTGATCATGCTGAGGAGGGGAGTGTGATGTTTCCACACC 902
QY 2480 oProProGly---LeuProAlaGly-----SerGlyProLeuAlaGlyProHisHis 2496

Db 903 CCCAGCGGGCGCTTCCCCCGGGGAGCGGGGCCCCCTCGGCTGGGCGCCCCAC 960

RESULT 13
BMS53310 1126 bp mRNA linear EST 20-FEB-2002
LOCUS AGENCOURT_6572642 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5467163
DEFINITION 5', mRNA sequence.
ACCESSION BMS53310
VERSION BMS53310.1 GI:18791930
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH-MGC http://mgi.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
TITLE Contact: Robert Strauberg, Ph.D.
COMMENT Email: cga@rs-remail.nih.gov
Tissue Procurement: DCTD/DRP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LHCMI968 row: 1 column: 12
High quality sequence stop: 674.
Location/Qualifiers
1. 1126
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5467163"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_1b="NIH MGC 41"
/note="Organ: skin, Vector: pOTB7, Site 1: XhoI, Site 2:
EcoRI, cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH-MGC Library."

ORIGIN

Alignment Scores:

Pred. No.:	3,18e-80	Length:	1126
Score:	1540.00	Matches:	314
Percent Similarity:	82.1%	Conservative:	9
Best Local Similarity:	79.90%	Mismatches:	16
Query Match:	11.65%	Indels:	55
DB:	4	Gaps:	1

US-09-522-753-5 (1-2517) x BMS53310 (1-1126)

QY 1457 LeuLysrThrApThrGlyAlaSerThrThGlySerIleYshIAspValArgSerLeu 1476

Db 2 CTCAGAGACAGACCGCGCGCTCCACACCTGCTCCAAAAGCAGCAGTACGCTCCCTC 61

QY 1477 IleGlySerProGlyArgThrPheProProValHisProLeuMetAlaAspAla 1496

Db 62 ATGGGAGCGCCCGCGAGCGTCCACCGGTGACCCGCTGATGTGATGGCCGAGCGCC 121

QY 1497 ArgAlaLeuGluIuArgAlaCyTYrGluGluSerLeuIleSerArgProGlyThrAlaSer 1516

Db 122 CGGAGCAGTGAAGCTGCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 181

QY 1517 SerSerGlyIleSerIleAlaArgGlyAlaProValIleValProGluLeuGlyIlePro 1536

Db 182 AGCTCGGGGGGCTCATTTGCGCGCGCGCGCGCTCATTTGCTGCTGAGTGGGCAAGCCG 241

QY 1537 ArgGlnSerProLeuThrTYrGluAspHisGlyValaProPheAlaGlyHisLeuProArg 1556

Db 242 CGGAGAGAGCGCCCTTAACCTATGAGAGACACGAGGAGACCTTTGCGGCGACCTCCACGA 301

QY 1557 GlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSer 1576

Db 302 GGTTCGCGCTGACAGCGGAGAGCCACCGCGCTGACAGAGAGGAGAGAGAGAGAGAG 361

QY 1577 SerIleValaSerGlnAspArgIleLeuThrSerThrProArgGlyIleAlaIleYserPro 1596

Db 362 AGCAGAGCATCCAGAGACCGAAAGCTGACCTGACCGCTGTGATGATCGCAAGTCCCGC 421

QY 1597 HisSerThrValProGluHisIleProHisProIleSerProTYrGluHisLeuLeuArg 1616

Db 422 CACAGCAGCTGCGCCGAGACACACACACACCTCTGCGCTTATGAGACCTGCTTCGG 481

QY 1617 GlyValSerGlyValaAspLeuTYrArgSerHisIleProLeuValaPheAspProThrSer 1636

Db 482 GCGGTGATGGCGTGAAGCTGTATCGGAGCCATCCCTGGCGCTTGCAGCCACCTCC 541

QY 1637 IleProArgGlyIleProLeuAspAlaAlaAlaIleTYrTYrLeuProArgHisIleVala 1656

Db 542 ATACCCCGCGCATCCCTTGAGCGCAGCGCTGCTACTACTGCTGCGGAGACCTGGCC 601

QY 1657 ProAspProThrTYrProHisIleuTYrProProTYrLeuIleArgGlyTYrProAspThr 1676

Db 602 CCCAGCCCGCATCCCGGACCTGTACCCACCTTACTCTCCGGGCTACCCGACAG 661

QY 1677 AlaAlaLeuGluIuAspArgGlnThrIleIleAspTYrIleThrSerGlnIleMetHis 1696

Db 662 GCGGCGGTGAGAAACCGGACAGCATCATCATCATCATCATCATCATCATCATCATCAT 721

QY 1697 HisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArg 1716

Db 722 CACAGCGCGGCGCACCGGACGAGCGCGAGCTGATGCTGAGGGGCTCTCGCCCGC 781

QY 1717 GluSerSerLeuAlaLeuAsnTYrAlaAlaGlyProArgGlyIleIleAspLeuSerGln 1736

Db 782 GAGTCTCTGCTGACATCACTACCTGCGGGTCCCGCA----- 820

QY 1737 ValProHisLeuProValLeuValProProThrProGlyThrProAlaThrAlaMetAsp 1756

Db 820 ----- 820

QY 1757 ArgLeuAlaTYrLeuProThrAlaProGlnProPheSerSerArgHisSerSerPro 1776

Db 820 ----- 820

QY 1777 LeuSerProGlyIleProThrHisLeuThrIleProThrThrThrSerSerGluArg 1796

Db 821 -----GAGAGTCCGAGACCTTGACAAACACACC-CCACGCTCTGTCGAAACGG 870

QY 1797 GluArgAspArg-AspArgGluArgAsp-ArgAspArgGluArgGlyIleLeuT 1816

Db 871 GAGCAGAGCCGGGAGATCAAGAGGAGGAGAACCGGATTCGAGCGGAAATCATCTCTCC 930

QY 1816 HisSerThrThrVal-GluHisIleProIleTyr-ArgProGlyIleThrGluInserSe 1835

Db 931 CGTCCACACCGAGCGTGGAGCGCCGACCCCTCTGGAACACCTGTAAACGAGCAACAA 990

QY 1835 rGlySerSerGlyIleSerSerGlyIleGlyIle 1845

Db 991 CGGAGCATCCGGGAAACTGGGGGGGGGGGT 1021

RESULT 14

BM477568 1057 bp mRNA linear EST 05-FEB-2002

LOCUS AGENCOURT_6484946 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5554420

DEFINITION 5', mRNA sequence.

ACCESSION BM477568

VERSION BM477568.1 GI:18526610

KEYWORDS EST.


```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5516745"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NH_MGC_41"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCACGAG(C). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NH_MGC library."

```

ORIGIN

Alignment Scores:

```

Pred. No.:      1.02e-76      Length:      1051
Score:          1479.50      Matches:      305
Percent Similarity: 78.95%      Conservative: 10
Best Local Similarity: 76.44%      Mismatches: 24
Query Match:    11.20%      Indels:      60
DB:             4           Gaps:      4

```

US-09-522-753-5 (1-2517) x BM423558 (1-1051)

```

QY      1457 LeuLeuYrYrApThrAlaSerThr-GlySerIlyeHisAspValArgSerLe 1476
DB      2 CTCAAGTACACATCCGGCGGCTCCACACATGGGCTCCAAAAAGACACACGTAACCTCTCT 61
QY      1476 uileGlySerProGlyYrGThrPheProProValHisProLeuAspValMetAlaAspAl 1496
DB      62 CATCGGAGCCCGGCGGAGAGTTCCACCCGCTCACCCGCTGATGTAGCGCGAGCC 121
QY      1496 aATgAlaLeuGluArgAlaCyTrYGluGluSerLeuIySerArgProGlyYThralase 1516
DB      122 CCGGGACACTGGAAGCTGCTGCTACAGAGAGAGCTGAGAGCGCGGACAGGACCGCCAG 181
QY      1516 rSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLyPr 1536
DB      182 CACTCGGGGGGCTCCATTGCGCGCGGCGCGCCCGCTGATGTGCTAGCTGGGCAAGCC 241
QY      1536 oATrGlnSerProLeuThrYrGluAspHisGlyAlaProPheAlaGlyHisLeuProAr 1556
DB      242 GCGGCAAGAGCCCTTAACCTATGAGACCAAGGGGCACTTTGCCGACCTCCACG 301
QY      1556 gGlySerProValThrMetArgGluProThrProArgLeuGlnGlySerLeuSerSe 1576
DB      302 AGGTTGCGCCGTGACCAAGCGGAGCCACAGCGCGCGCTGAGAGGAGGCAAGCTTCGTC 361
QY      1576 rSerIyAlaSerGlnAspArgIyLeuThrSerThrProArgGluIleAlaIySerPr 1596
DB      362 CACCAAGGACATCCCAAGACCGAAAGCTGAGCTGACCTCTGTGAGATCCCAAGTCCCC 421
QY      1596 oHisSerThrValProGluHisAspProHisAspProIleSerProGlyGluHisLeuLeuAr 1616
DB      422 GCAAGAGACCTGCGCGAGACCAACCAACCCCATCTGCGCTTATGAGACCTGCTTCG 481
QY      1616 gGlyValSerGlyValAspLeuYrArgSerHisIleProLeuAlaPheAspProThrSe 1636
DB      482 GGGCGTGAAGTGGCGTGAACCTGTATGCAAGCAATCCCGCTGACCTTCGACCCCACTC 541
QY      1636 rIleProArgGlyIleProLeuAspAlaAlaAlaIleYrTyrLeuProArgHisLeuAl 1656
DB      542 CATATCCCGGGGATCTCTGTGACGAGCGCGCTGCTACTACCTGCGCGACACCTGGC 601
QY      1656 aProAsnProThrTyrProHisLeuYrProProYrLeuIleArgGlyTyrProAspTh 1676
DB      602 CCCCAACCCCACTACCTGACCTGTACCACTTACTCATCCGCGCTAACCCCAAC 661
QY      1676 rAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspYrIleThrSerGlnGlnMetH 1696

```

```

DB      662 GCGCGCGCTGAGAACCGGACGACCATCATCATATGACTACCTCGACGAGATGCA 721
QY      1696 rHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProAr 1716
DB      722 CCACAAAGCGGACCGCCCATGCGCCAGCGAGCTGATATGCTGAAGGGGCTCTGCGCCG 781
QY      1716 gGlySerSerLeuAlaLeuAsnYrAlaAlaGlyProArgGlyIleIleAspLeuSerG 1736
DB      782 CGAGTCTCGCTGCACTCAACTACGCTGGGGGTCCC----- 818
QY      1736 nValProHisLeuProValLeuValProProThrProGlyThrProAlaThrAlaMetAs 1756
DB      818 ----- 818
QY      1756 pArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSerPr 1776
DB      818 ----- 818
QY      1776 oLeuSerPro-GlyGlyProThrHis-LeuThrLyAspProThrThrThrSerSerSerGlu 1795
DB      819 -----CCAGGAGAGGCCAACCCACCTTACCAACCAACCCCAAGCTCCCTCGAG 871
QY      1796 ArgGluArgAspArg--AspArgGluArgAspArgAspArgGlu-ArgGluLySerI 1814
DB      872 CCGGAGCGGAGAACCGGGATCCAAAGCGCGGACCGGGATCGGGAAACCGGAAAGGTCCAT 931
QY      1814 eleuThr---SerThrThrThrValGlnHisAlaProIleTrp-ArgProGlyYThrGlnG 1833
DB      932 CTTTCCTTCCCCCCCAACGAGGGAGAGCAAGCCCCCATCTGGAACCTGTGTAACCAAC 991
QY      1833 InSerSerGlySerSerGlySer-----SerGlyGlyGly 1845
DB      992 CCAACAACGCGGACCAAGGSCACTTACACTTCCGGGGGGGGGGG 1038

```

Search completed: November 12, 2005, 03:26:38
 Job time : 24452 secs

THIS PAGE BLANK (USPTO)